AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1

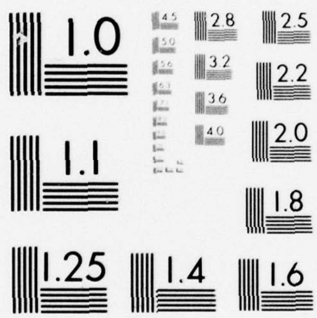
APPROXIMATION THEORY II

NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.

UNCLASSIFIED AFOSR TR-76-1242 N/L

1 OF 7

AD A033032



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TRIP TIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-76-1242	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APPROXIMATION THEORY II		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. G. Lorentz, C. K. Chui, L. L. Schumaker		8. CONTRACT OR GRANT NUMBER(s) AFOSR 76-2921
9. PERFORMING ORGANIZATION NAME AND ADDRESS The University of Texas at Austin, Department of Mathematics, Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A3
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NM) Bldg 410, Bolling AFB DC 20332		12. REPORT DATE November 1976
		13. NUMBER OF PAGES 588
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Academic Press, Inc., New York 1976

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This book contains the proceedings of a Symposium on Approximation Theory which was held in Austin on January 18-21, 1976. The meeting was attended by more than 120 mathematicians. In addition to the United States, the other countries represented were Belgium, Brazil, Canada, France, German Federal Republic, Hungary, Israel, Italy, Netherlands, New Zeland, Sweden, and the USSR. The organizing committee for the symposium consisted of E. W. Cheney, C. K. Chui, G. G. Lorentz(director), and L. L. Schumaker.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A033032

$ 23.50

DR

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Approximation Theory II

ACADEMIC PRESS RAPID MANUSCRIPT REPRODUCTION

Proceedings of an International Symposium Conducted by The University of Texas at Austin, Texas, January 18-21, 1976

(1)

Approximation Theory II

Edited by

G. G. Lorentz

Department of Mathematics
University of Texas
Austin, Texas

C. K. Chui

Department of Mathematics
Texas A&M University
College Station, Texas

L. L. Schumaker

Department of Mathematics
University of Texas
Austin, Texas

ACADEMIC PRESS, INC.
NEW YORK SAN FRANCISCO LONDON 1976
A Subsidiary of Harcourt Brace Jovanovich, Publishers

iii

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 73-2083

ISBN 0-12-455760-0

PRINTED IN THE UNITED STATES OF AMERICA

CONTENTS

CONTENTS

CONTENTS

LIST OF CONTRIBUTORS

R. B. Barrar, Department of Mathematics, University of Oregon, Eugene, Oregon 97403

H. Berens, Mathem. Institut der Universität, Bismarckstr. 1 1/2, 582 Erlangen, FEDERAL REPUBLIC OF GERMANY

Jörg Blatter, Instituto de Mathematica Pura e Aplicada, Rua Luiz de Camoes 68, Rio de Janeiro, BRAZIL

Carl de Boor, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706

Dietrich Braess, Institut für Mathematik der Universität, 463 Bochum, FEDERAL REPUBLIC OF GERMANY

R. Creighton Buck, Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706

H. G. Burchard, Department of Mathematics and Statistics, Oklahoma State University, Stillwater, Oklahoma 74074

P. L. Butzer, Lehrstuhl A für Mathematik, Technological University of Aachen, Tempelgraben 55, 51 Aachen, FEDERAL REPUBLIC OF GERMANY

B. L. Chalmers, Department of Mathematics, University of California, Riverside, California 92507

E. W. Cheney, Department of Mathematics, RLM 8-100, The University of Texas, Austin, Texas 78712

Charles K. Chui, Department of Mathematics, Texas A&M University, College Station, Texas 77843

L. Collatz, Institut für Agnew. Mathematik der Universität, Bundesstr. 55, 2 Hamburg, FEDERAL REPUBLIC OF GERMANY

Ronald A. DeVore, Department of Mathematics, Oakland University, Rochester, Michigan 48063

Z. Ditzian, Department of Mathematics, University of Alberta, Alberta, CANADA

P. Erdös, Mathematics Institute, Academy of Sciences, Budapest, HUNGARY

Stephen D. Fisher, Department of Mathematics, Northwestern University, Evanston, Illinois 60201

C. Franchetti, Via Si Türr 4, 50137 Firenze, ITALY

Geza Freud, Department of Mathematics, Ohio State University, Columbus, Ohio 43210

K.-H. Hoffmann, III Mathem. Institut der Freien Universität, Arnimallee 2-6, 1 WEST BERLIN 33

R. B. Holmes, Department of Mathematics, Purdue University, Lafayette, Indiana 47907

L. Jaffe, Department of Mathematics, RLM 8-100, The University of Texas, Austin, Texas 78712

K. Jetter, Mathematisches Institut, Universität Tübingen, 74 Tübingen, FEDERAL REPUBLIC OF GERMANY

Darell J. Johnson, Department of Mathematics, New Mexico State University, Las Cruces, New Mexico 88001

A. Klostermair, III Mathem. Institut der Freien Universität, Arnimallee 2-6, 1 WEST BERLIN 33

D. Leviatan, Department of Mathematics, Tel-Aviv University, Tel-Aviv, ISRAEL

H. L. Loeb, Department of Mathematics, University of Oregon, Eugene, Oregon 97403

G. G. Lorentz, Department of Mathematics, RLM 8-100, The University of Texas, Austin, Texas 78712

W. A. J. Luxemburg, Department of Mathematics, California Institute of Technology, Pasadena, California 91125

Lois Mansfield, Department of Computer Sciences, University of Kansas, Lawrence, Kansas 66044

M. J. Marsden, Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Preceding page blank

LIST OF CONTRIBUTORS

Avraham A. Melkman, Department of Mathematics, Hebrew University, Jerusalem, ISRAEL
F. T. Metcalf, Department of Mathematics, University of California, Riverside, California 92507
Charles A. Micchelli, Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598
R. J. Nessel, Lehrstuhl A für Mathematik, Technological University of Aachen, Tempelgraben 55, 51 Aachen, FEDERAL REPUBLIC OF GERMANY
G. P. Nevai, 12 rue Clairaut, 75017 Paris, FRANCE
J. T. Oden, Department of Aer. Eng.& Eng. Mech., The University of Texas, Austin, Texas 78712
Charles F. Osgood, Naval Research Laboratory, Math. Res. Ctr., Code 7840, Bldg. 43, Washington, D. C. 20390
Eli Passow, Department of Mathematics, Temple University, Philadelphia, Pennsylvania 19122
A. Pinkus, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706
S. D. Riemenschneider, Department of Mathematics, University of Alberta, Edmonton, CANADA
T. J. Rivlin, Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
John A. Roulier, Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27607
E. B. Saff, Department of Mathematics, University of Southern Florida, Tampa, Florida 33620
R. Schaback, Institut für Angewandte Mathematik, Universität Bonn, 53 Bonn, FEDERAL REPUBLIC OF GERMANY
A. H. Schatz, Department of Mathematics, Cornell University, Ithaca, New York 14850
Karl Scherer, Institut für Angewandte Mathematik, Universität Bonn, 53 Bonn, FEDERAL REPUBLIC OF GERMANY
Larry L. Schumaker, Department of Mathematics, RLM 8-100, The University of Texas, Austin, Texas 78712
A. Sharma, Department of Mathematics, University of Alberta, Edmonton, CANADA
Oved Shisha, Department of Mathematics, University of Rhode Island, Kingston, Rhode Island 02881
Philip W. Smith, Department of Mathematics, Texas A&M University, College Station, Texas 77843
R. L. Stens, Lehrstuhl A für Mathematik, Technological University of Aachen, Tempelgraben 55, 51 Aachen, FEDERAL REPUBLIC OF GERMANY
S. S. Stangler, Mathematisches Institut, Universität Tübingen, 74 Tübingen, FEDERAL REPUBLIC OF GERMANY
J. Tzimbalario, Department of Mathematics, University of Alberta, Edmonton, CANADA
R. S. Varga, Department of Mathematics, Kent State University, Kent, Ohio 44240
A. G. Vituškin, Steklov Mathematical Institute, Vavilova 42, Moscow-333, USSR
L. B. Wahlbin, Department of Mathematics, Cornell University, Ithaca, New York 14850
H. Werner, Rechenzentrum, Universität Münster, Roxelerstr. 60, 44 Münster, FEDERAL REPUBLIC OF GERMANY
G. Wilmes, Lehrstuhl A für Mathematik, Technological University of Aachen, Tempelgraben 55, 51 Aachen, FEDERAL REPUBLIC OF GERMANY
D. Wulbert, Department of Mathematics, University of California, La Jolla, California 92037
M. Zamansky, Department of Mathematics, University of Paris, Paris, FRANCE
K. L. Zeller, Mathematisches Institut, Universität Tübingen, 74 Tübingen, FEDERAL REPUBLIC OF GERMANY

PREFACE

This book contains the proceedings of a Symposium on Approximation Theory which was held in Austin on January 18-21, 1976. The meeting was attended by more than 120 mathematicians. In addition to the United States, the other countries represented were Belgium, Brazil, Canada, France, German Federal Republic, Hungary, Israel, Italy, Netherlands, New Zealand, Sweden, and the USSR. The organizing committee for the symposium consisted of E.W. Cheney, C.K. Chui, G.G. Lorentz (director), and L.L. Schumaker.

The program of the symposium included seven one-hour invited expository lectures, more than forty one-half-hour invited research talks, and a large number of contributed fifteen minute talks. Because of space limitations, we have been able to publish only the invited papers.

The expository papers describe recent developments in some important special fields. For example, D. Braess discusses nonlinear approximation, C. de Boor surveys the central role of *B*-splines in the general theory of splines. "This exhibits the essentially local, but not completely local character of splines." R. DeVore discusses the degree of approximation by polynomials, trigonometric polynomials, and splines, obtaining the results from a relation between the moduli of smoothness and Peetre's *K*-functionals. A. G. Vituškin (from the Steklov Mathematical Institute in Moscow) has a new point of view upon entropy; he indicates its applications to coding. G. P. Nevai describes recent results (by himself, G. Freud and others) in the theory of convergence of interpolation formulas and of orthogonal polynomials. C. K. Chui surveys Padé approximations of arbitrary analytic functions. L. L. Schumaker discusses fitting surfaces; it is hoped that this exposition will be useful for practical computation. The shorter research papers cover a wide variety of topics from the extremal properties of polynomials by P. Erdös to approximation of fixed points by L. Collatz. Several papers are devoted to spline theory and to the Birkhoff interpolation problem.

Our thanks are due to the National Science Foundation and the Air Force Office of Scientific Research for grants which made the symposium possible. The latter organization provided support for applied papers (Numerical Analysis, Finite Element Methods). We would also like to thank the University of Texas at Austin for providing secretarial help and the facilities of the Thompson Conference Center.

C.K. Chui, G.G. Lorentz, L.L. Schumaker

SPLINES AS LINEAR COMBINATIONS OF B-SPLINES. A SURVEY

Carl de Boor

This paper is intended to serve as a postscript to the fundamental 1966 paper by Curry and Schoenberg on B-splines. It is also intended to promote the point of view that B-splines are truly basic splines: B-splines express the essentially local, but not completely local, character of splines; certain facts about splines take on their most striking form when put into B-spline terms, and many theorems about splines are most easily proved with the aid of B-splines; the computational determination of a specific spline from some information about it is usually facilitated when B-splines are used in its construction.

1. Introduction

The layout of the survey is as follows. After a short discussion of cardinal B-splines, i.e., of B-splines on a uniform knot sequence, in Section 2, B-splines for an arbitrary knot sequence are introduced in Section 3 and shown to be a basis for certain spaces of piecewise polynomial functions. Various simple properties of B-splines are listed in Section 4, and the relationship between a spline and its coordinates with respect to a B-spline basis is explored in Section 5. This leads naturally into the discussion of local spline approximation schemes, in Section 6. Results concerning existence and uniqueness of interpolating splines and the related total positivity and variation diminishing properties of B-splines are presented in Section 7. Section 8 describes the connection between splines and certain "best" interpolation schemes. Finally, Section 9 is devoted to generalized B-splines and ends with a new definition of polynomial B-splines in many variables due to I. J. Schoenberg.

No claim of completeness is made, and the author would be grateful to hear of any omissions.

The following notation is used throughout the paper, usually without further explanation:

$\mathbb{Z}$ denotes the set of integers, $\mathbb{R}$ the set of real numbers, and A^B the set of functions on B into A. Thus, $\mathbb{R}^{\mathbb{Z}}$ is the set of real bi-infinite sequences.

$m(B)$ is the linear space of bounded real functions on B, normed by $\|f\|_{\infty,B} := \sup_{x\in B}|f(x)|$. For $1 \le p \le \infty$, $\mathbb{L}_p(I)$ denotes the space of (equivalence classes of) functions f on the interval I for which $\|f\|_p := \|f\|_{p,I} := (\int_I |f|^p)^{1/p} < \infty$. $C^k(I)$ is the space of k times continuously differentiable functions on I. $\mathbb{L}_p^k(I)$ is the subspace of those $f \in C^{k-1}(I)$ whose $(k-1)$st derivative is absolutely continuous and whose k^{th} derivative is in $\mathbb{L}_p(I)$. $\mathbb{M}^k(I)$ is the subspace of $C^{k-2}(I)$ whose elements have an absolutely continuous $(k-2)^{nd}$ derivative and a $(k-1)^{st}$ derivative of bounded variation. Finally, $\ell_p(\mathbb{Z}) := \{\alpha \in \mathbb{R}^{\mathbb{Z}} \mid \|\alpha\|_p := (\sum_i |\alpha_i|^p)^{1/p} < \infty\}$.

$\mathbb{P}_k$ denotes the linear space of all polynomials of order k (or, degree $< k$) with real coefficients. For a strictly increasing sequence $\underline{\underline{\xi}} := (\xi_i)$, $\mathbb{P}_{k,\underline{\underline{\xi}}}$ denotes the linear space of all piecewise polynomial (or, pp) functions of order k on $I := [\inf \xi_i, \sup \xi_i]$ with breakpoint sequence $\underline{\underline{\xi}}$. Explicitly, $f \in \mathbb{P}_{k,\underline{\underline{\xi}}}$ iff $f\big|_{(\xi_i,\xi_{i+1})} \in \mathbb{P}_k\big|_{(\xi_i,\xi_{i+1})}$, all i. In addition, $f \in \mathbb{P}_{k,\underline{\underline{\xi}}}$ is taken to have two values at ξ_i, i.e., the values $f(\xi_i^-)$ and $f(\xi_i^+)$. If the reader finds it necessary to think of f as a single-valued function, he should choose some rule $f(\xi_i) := \alpha f(\xi_i^-) + (1-\alpha)f(\xi_i^+)$ (e.g., $\alpha = 1/2$) and stick with it.

If $\underline{\underline{v}} = (v_i)$ is a sequence of nonnegative integers

corresponding to $\underline{\underline{\xi}}$, then $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$ denotes the linear subspace of $\mathbb{P}_{k,\underline{\underline{\xi}}}$ consisting of those $f \in \mathbb{P}_{k,\underline{\underline{\xi}}}$ for which

$$\text{jump}_{\xi_i} f^{(v)} = 0 \quad \text{for} \quad v < v_i, \quad \text{all} \quad i.$$

The v-th derivative of f is also denoted by $D^v f$ as well as by $f^{(v)}$. $[\tau_0,\ldots,\tau_k]f$ stands for the k-th divided difference of f at the points $\tau_0,\ldots,\tau_k$. In particular, $[\tau_0]f = f(\tau_0)$.

$\text{const}_{\alpha,\ldots,\omega}$ denotes a constant which may depend on the quantities $\alpha,\ldots,\omega$.

2. Cardinal Splines

B-splines made their first appearance in Schoenberg's 1946 paper on the approximation of equidistant data by analytic functions. There is no doubt that B-splines appear in earlier literature. They play a prominent role already in Favard's work [35], and Schoenberg has always maintained that they were already known to Laplace (see [70, p. 68]). But it is in Schoenberg's paper that they were thought important enough to be given a name, "basic k^{th}-order spline curves." Since this is the same paper in which Schoenberg introduces splines, I happily conclude that B-splines were there at the very beginning.

Schoenberg introduces the B-spline, <u>neé</u> basic spline curve, <u>alias</u> spline frequency function [29] <u>alias</u> fundamental spline function [71, 30]

$$(2.1) \quad M_k(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\sin u/2}{u/2}\right)^k e^{iux}\, du$$

and then observes that

$$(2.2) \quad M_k(x) = k[-k/2, 1-k/2, \ldots, k/2](\cdot - x)_+^{k-1},$$

i.e., $M_k(x)$ is k times the k-th divided difference in y at

the $k+1$ points $j-k/2$, $j=0,\ldots,k$, of the function $(y-x)_+^{k-1} := (\max\{0,y-x\})^{k-1}$. These formulae show that M_k is the k-th convolution power of the characteristic function of the interval $[-1/2,1/2]$,

$$
M_1(x) = \begin{cases} 1, & \text{for } x \in (-1/2,1/2) \\ 0, & \text{for } x \notin [-1/2,1/2] \end{cases},
$$

(2.3)

$$
M_k(x) = (M_i * M_j)(x) = \int_{-\infty}^{\infty} M_i(x-y)M_j(y)\,dy \quad \text{for } i+j=k.
$$

Therefore,--and this is why Laplace must have known B-splines-- M_k is the density distribution of the error committed on the sum of k independent real random variables if each variable is replaced by its nearest integer value [70, p. 76].

It is easily seen from (2.2) or (2.3) that

$$
M_k \in \mathbb{P}_{k,\mathbb{Z}+k/2} \cap C^{k-2} =: \text{ set of "spline curves of order k"}
$$

as Schoenberg calls them. The subject matter of the paper [70] is the study of approximations of the form

$$
Af := \sum_{n\in\mathbb{Z}} f(n)L(\cdot - n),
$$

and the B-splines come in because they offer a convenient way of expressing, and thereby analyzing, the various pp "basic" functions L considered in the paper.

In the 60's, Schoenberg's results were rediscovered and considerably extended by those engaged in studying the mathematical aspects of the finite element method (see Aubin [1,2], Babuška [3], Bramble and Hilbert [19], Fix and Strang [38] and Strang and Fix [80], Di Guglielmo [33], and others). When restricted to the one-dimensional setting of Schoenberg's paper, these people are seen to consider approximation processes of the form

$$Af := \sum_{n\in\mathbb{Z}} (\lambda f(\cdot +n))L(\cdot -n)$$

for some convenient basic function L, e.g., $L = M_k$, and some linear functional $\lambda \in C^*(\mathbb{R})$, and to study the convergence behavior of

$$A_h := S_{1/h}AS_h, \qquad \text{with} \quad (S_\alpha f)(x) := f(\alpha x),$$

as $h \to 0$. The results of this study are nicely summarized by Link [57].

Schoenberg himself developed a particular aspect of his '46 paper, viz. Cardinal spline interpolation, in considerable detail in a sequence of seven papers which appeared in the late 60's and early 70's. These papers have become the basis for his beautiful monograph [76] on cardinal spline interpolation. Readers interested in the properties and use of B-splines on <u>uniform</u> knot sequences are urged to consult that monograph.

3. <u>B-splines Defined</u>

It was apparently Schoenberg's colleague H. B. Curry who observed that the formulation (2.2) of M_k as a k-th order difference generalizes naturally to a k-th order divided difference on arbitrary points,

$$(3.1) \quad M_{i,k}(x) := k[t_i,\ldots,t_{i+k}](\cdot -x)_+^{k-1}.$$

The resulting paper [30], though written in 1946 (see [29]), was finally published in 1966. The function $M_{i,k}$ is easily seen to be a pp function of order k with breakpoints $t_i,\ldots,t_{i+k}$, and with smoothness across each breakpoint t_j which depends on its multiplicity, i.e., on the frequency with which the number t_j occurs in the sequence $t_i,\ldots,t_{i+k}$. Further, one readily sees that

$$(3.2) \quad M_{i,k}(x) \geq 0 \quad \text{with equality if} \quad x \notin (t_i,t_{i+k})$$

in case $t_i \le \dots \le t_{i+k}$.

Now let $\underline{\underline{t}} := (t_i)_{-\infty}^{\infty}$ be nondecreasing, with

$$t_{-\infty} := \inf t_i, \; t_{\infty} := \sup t_i,$$

and let $(M_{i,k})_{-\infty}^{\infty}$ be the corresponding B-spline sequence. Then, the prescription

$$(\Sigma_i \alpha_i M_{i,k})(x) := \Sigma_i \alpha_i M_{i,k}(x), \quad \text{i.e., pointwise,}$$

makes sense for all $x \in \mathbb{R}$ and all $\alpha \in \mathbb{R}^{\mathbb{Z}}$ since, by (3.2), at most k of the terms in the second sum are nonzero for any given x.

In a later publication [73] (but see already Curry's review [28] of [70]), Schoenberg gave these functions $M_{i,k}$ the name <u>basic spline</u>, or <u>B-spline</u>, for the following reason.

THEOREM 3.1 [30]. <u>If</u> $\underline{\underline{t}} := (t_i)_{-\infty}^{\infty}$ <u>is nondecreasing, with</u> $t_i < t_{i+k}$ <u>and</u> $d_i := \text{card}\{j \mid t_j = t_i\}$, <u>all</u> i, <u>then the corresponding sequence</u> $(M_{i,k})_{-\infty}^{\infty}$ <u>of B-splines is a basis for the linear space</u> $\mathcal{S}_{k,\underline{\underline{t}}}$ <u>of all functions</u> f <u>on</u> $\mathbb{R}$ <u>which vanish off</u> $(t_{-\infty}, t_{\infty})$ <u>and which, on</u> $(t_{-\infty}, t_{\infty})$, <u>satisfy</u>

$$f\Big|_{(t_i,t_{i+1})} \in \mathbb{P}_k\Big|_{(t_i,t_{i+1})}, \quad \text{jump}_{t_i} f^{(r)} = 0 \;\; \underline{\text{for}} \;\; r < k - d_i,$$

<u>all</u> i,

<u>in the sense that the map</u> $\mathbb{R}^{\mathbb{Z}} \to \mathcal{S}_{k,\underline{\underline{t}}} : \alpha \mapsto \Sigma_i \alpha_i M_{i,k}$ <u>is one-one and onto</u>.

This theorem motivates the definition

$$\mathcal{S}_{k,\underline{\underline{t}}} := \{\Sigma_i \alpha_i M_{i,k} \mid \alpha_i \in \mathbb{R}, \text{ all } i\}$$

for <u>arbitrary</u> nondecreasing $\underline{\underline{t}}$, bi-infinite or not, with the sum taken over all i for which $(t_i, \dots, t_{i+k})$ is a segment of $\underline{\underline{t}}$. In particular,

$$\text{if } \underline{\underline{t}} = (t_i)_1^{n+k}, \text{ then } \mathcal{S}_{k,\underline{\underline{t}}} = \left\{ \sum_{i=1}^{n} \alpha_i M_{i,k} \mid \alpha \in \mathbb{R}^n \right\}.$$

Further, we will call $\mathcal{S}_{k,\underline{\underline{t}}}$ the collection of (polynomial) <u>splines of order</u> k <u>with knot sequence</u> $\underline{\underline{t}}$.

COROLLARY (Construction of a B-spline basis for $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$). <u>Let</u> $\underline{\underline{\xi}} := (\xi_i)_1^{p+1}$ <u>be strictly increasing,</u> $\underline{\underline{v}} := (v_i)_1^{p+1}$ <u>be a corresponding sequence of integers in</u> $[0,k]$ <u>with</u> $v_1 = v_{p+1} = 0$, <u>and let</u> $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$ <u>be the space of</u> pp <u>functions of order</u> k <u>on</u> $[\xi_1,\xi_{p+1}]$ <u>with breakpoints</u> $\xi_2,\ldots,\xi_p$ <u>and continuous v-</u>th <u>derivative at</u> ξ_i <u>for</u> $v < v_i$, <u>all</u> i. <u>If</u>

$$\underline{\underline{t}} := (t_i)_1^{n+k} = (\underbrace{\xi_1,\ldots,\xi_1}_{v_1=k}, \underbrace{\xi_2,\ldots,\xi_2}_{v_2},\ldots, \underbrace{\xi_{p+1},\ldots,\xi_{p+1}}_{v_{p+1}=k}),$$

<u>then</u> $n = k + \Sigma_2^p(k-v_i)$ <u>and the sequence</u> $(M_{i,k})_1^n$ <u>of B-splines (restricted to</u> $[\xi_1,\xi_{p+1}]$<u>)</u> <u>of order</u> k <u>for the knot sequence</u> $\underline{\underline{t}}$ <u>is a basis for</u> $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$.

For k even, $k = 2m$, it is customary to single out the subspace S of so-called "natural" splines in $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$. This subspace consists of those f in $\mathbb{P}_{k,\underline{\underline{\xi}},\underline{\underline{v}}}$ for which $f\big|_{(\xi_1,\xi_2)}$ and $f\big|_{(\xi_p,\xi_{p+1})}$ are both of degree $< m$ (see Section 8 below (8.8)). Greville [44] has described the following B-spline like basis for S,

$$\hat{M}_{m+1,k},\ldots,\hat{M}_{k,k},M_{k+1,k},\ldots,M_{n-k,k},\hat{M}_{n-k+1,k},\ldots,\hat{M}_{n-m,k}$$

with the special functions $\hat{M}_{i,k}$ defined as follows:

$$(3.3)\quad \begin{aligned} \hat{M}_{i,k}(x) &:= k[t_{k+1},\ldots,t_{i+k}](\cdot - x)_+^{k-1} \quad \text{for } i \le k \\ \hat{M}_{n-i,k}(x) &:= (-1)^k k[t_{n-i},\ldots,t_n](x-\cdot)_+^{k-1} \quad \text{for } i < k. \end{aligned}$$

For a different generalization of M_k to a "B-spline" with multiple knots (which are otherwise uniformly spaced), see Schoenberg and Sharma [77] and Lecture 5 of Schoenberg's monograph [76]. Certain technical assumptions made by them in their construction have recently been removed by Lee [56].

4. Simple Properties of the B-spline

In this section, we list some simple properties of the B-spline, some of which are enlarged upon in subsequent sections. The definition of $M_{i,k}$ as a divided difference together with Taylor's formula with integral remainder readily imply that, for $t_i < t_{i+k}$,

$$(4.1)\quad [t_i,\ldots,t_{i+k}]f = \int M_{i,k}(s)\, f^{(k)}(s)\, ds/k!, \qquad \text{all } f \in \mathbb{L}_1^k[t_i,t_{i+k}].$$

In particular,

$$(4.2)\quad \int M_{i,k}(s)\, ds = 1.$$

This shows that, on $[t_i,t_{i+k}]$,

$$(4.3)\quad \phi(x) := \int_{-\infty}^{x} M_{i,k}(s)\, ds$$

is a spline of order $k+1$ with knots $t_i,\ldots,t_{i+k}$ and rises strictly monotonely from a value of 0 at t_i (and to left of t_i) to a value of 1 at t_{i+k} (and to the right of t_{i+k}). This function is therefore useful in constructing piecewise monotone spline interpolants as is done in Passow [66], but without having to resort to multiple knots as he does. One

obtains his construction as a special case by letting half the t_j's equal t_i and the other half equal t_{i+k}. Use of ϕ also produces a very quick proof that splines have property SAIN with respect to interpolation at a given set of points and the uniform norm (see Chui, Rozema, Smith, and Ward [24], who use (4.3) in the form (4.11)). Because of its local and monotone character, ϕ has also been instrumental in DeVore's successful investigation [32] of the order of approximation to smooth monotone functions by monotone splines.

It seems more convenient in computations to use the <u>normalized</u> B-spline

$$(4.4)\quad N_{i,k}(x) := ([t_{i+1},\dots,t_{i+k}] - [t_i,\dots,t_{i+k-1}])(\cdot - x)_+^{k-1}$$
$$= (t_{i+k} - t_i)M_{i,k}(x)/k,$$

since it insures (see (5.8) below) that

$$(4.5)\quad \sum_{i=1}^{n} N_{i,k} = 1 \quad \text{on} \quad [t_k, t_{n+1}].$$

Note that then

$$(4.6)\quad N_{i,k}^{(1)} = M_{i,k-1} - M_{i+1,k-1}$$
$$= \frac{k-1}{t_{i+k-1} - t_i} N_{i,k-1} - \frac{k-1}{t_{i+k} - t_{i+1}} N_{i+1,k-1}.$$

If one follows [8] and applies Leibniz' formula

$$(4.7)\quad [t_i,\dots,t_j](fg) = \sum_{r=i}^{j} [t_i,\dots,t_r]f\ [t_r,\dots,t_j]g$$

for the divided difference of a product to

$$(\cdot - x)_+^{j} = (\cdot - x)_+^{j-1}(\cdot - x)$$

and notes that all divided differences of $(\cdot - x)$ of order > 1

vanish, then one obtains the recurrence relation

$$(4.8) \quad [t_i,\ldots,t_{i+k}](\cdot-x)_+^j =$$

$$\left(\frac{x-t_i}{t_{i+k}-t_i}[t_i,\ldots,t_{i+k-1}] + \frac{t_{i+k}-x}{t_{i+k}-t_i}[t_{i+1},\ldots,t_{i+k}]\right)(\cdot-x)_+^{j-1},$$

which in turn implies that

$$(4.9) \quad \frac{k-j-1}{k-1}N_{i,k}^{(j)}(x) = \frac{x-t_i}{t_{i+k-1}-t_i}N_{i,k-1}^{(j)}(x) + \frac{t_{i+k}-x}{t_{i+k}-t_{i+1}}N_{i+1,k-1}^{(j)}(x).$$

For $j = 0$, this recurrence was found by the author [8] and by L. Mansfield, and by Cox [27] who proved it by a different argument and for distinct knots only, and gave a backward error analysis in that case for the evaluation algorithm based on the recurrence. The recurrence provides a scheme for the stable evaluation of B-splines since, on the interval (t_i,t_{i+k}) of interest, i.e., on the support of $N_{i,k}$, both weights in (4.9) are positive. This observation also allows us to establish, by induction on k, that

$$(4.10) \quad N_{i,k} > 0 \quad \text{on} \quad (t_i,t_{i+k}).$$

Similar recurrence relations for the integral of a B-spline have been given by Gaffney [39], and for the integral of products of B-splines by Lyche, Schumaker, and the author [17]. In this connection, we note that

$$(4.11) \quad \int_{-\infty}^{x} M_{i,k}(s)\,ds = \sum_{j=i}^{i+r} N_{j,k+1}(x) \quad \text{for} \quad x \le t_{i+r+1}.$$

B-splines are convenient for relating splines with multiple knots to splines with simple knots and vice versa (e.g., [7], Rice [68], Burchard [23], also the paper by P. Smith in these proceedings), since a B-spline <u>is a continuous function of its knots</u>, within reason. Specifically, writing

$$N_{t_i,\dots,t_{i+k}} := N_{i,k}$$

to stress the dependence of $N_{i,k}$ on its knots $t_i,\dots,t_{i+k}$, <u>the map</u>

$$(\tau_j)_0^k \mapsto N_{\tau_0,\dots,\tau_k}$$

<u>is continuous as a map from</u> $\{\tau \in \mathbb{R}^{k+1} | \tau_0 \le \dots \le \tau_k, \tau_j < \tau_{j+k-1}\}$ <u>to</u> $C(\mathbb{R})$; <u>it is also continuous as a map from</u> $\{\tau \in \mathbb{R}^{k+1} | \tau_0 \le \dots \le \tau_k, \tau_0 < \tau_k\}$ <u>to</u> $\mathbb{L}_p(\mathbb{R})$ <u>for every</u> $1 \le p < \infty$.

The precise behavior of $N_{i,k}$ near the boundary of its support can be read off directly from its definition as a divided difference. Since

$$(y-x)_+^{k-1} - (-)^k(x-y)_+^{k-1} = (y-x)^{k-1},$$

and the k-th divided difference of a polynomial of order k vanishes, one can write $N_{i,k}$ also in the form

$$N_{i,k}(x) = (-)^k([t_{i+1},\dots,t_{i+k}] - [t_i,\dots,t_{i+k-1}])(x-\cdot)_+^{k-1}. \tag{4.12}$$

From this, one infers at once that, e.g., for x near t_i,

$$N_{i,k}(x) = (x-t_i)_+^{k-r} \prod_{j=1}^{k-r} \frac{k-j}{t_{i+k-j}-t_i} + O((x-t_i)_+^{k-r+1}) \tag{4.13}$$

if $t_i = t_{i+r-1} < t_{i+r}$, hence

$$N_{j,k}(x) = \begin{array}{ll} O((x-t_i)N_{i,k}(x)) & \text{for } j > i \text{ as } x \to t_i \\ O((x-t_{i+k})N_{i,k}(x)) & \text{for } j < i \text{ as } x \to t_{i+k}. \end{array} \tag{4.14}$$

If $t_i < \dots < t_{i+k}$, then $N_{i,k}$ has a zero of order k-1 at t_i by (4.13) and also a k-1 fold zero at t_{i+k} by symmetry. This implies that

$$(4.15)\quad 0 = N_{i,k}^{(j-r)}(t_{j+k}) = \int_{t_i}^{t_{i+k}} (t_{i+k}-s)^{r-1} N_{i,k}^{(j)}(s)\,ds/(r-1)!,$$
$$r = 1,\ldots,j;\ j = 1,\ldots,k-1,$$

showing that $N_{i,k}^{(j)}$ <u>is orthogonal to</u> $\mathbb{P}_j$ <u>on</u> $[t_i, t_{i+k}]$, $j = 1, \ldots, k-1$. (This fact was pointed out to me in 1973 by H. G. Burchard.)

5. The B-spline Series

In this section the relationship

$$(5.1)\quad \Sigma_i \alpha_i N_i \leftrightarrow \alpha$$

between a spline and the sequence of its B-spline coefficients (with respect to the normalized B-splines) is discussed. Further aspects of this relationship will be mentioned in subsequent sections. From here on, we suppress the subscript k in $N_{i,k}$ and $M_{i,k}$ except when necessary. *Also, we restrict* the knot sequence $\underline{\underline{t}}$ to be bi-infinite in order to avoid (mostly notational) complications. This is no essential restriction since any spline can always be extended to a spline with a bi-infinite knot sequence merely by adding to its expansion appropriate B-splines with zero coefficients.

A B-spline series may be differentiated by differencing the coefficients. Precisely, repeated application of (4.6) gives

$$(5.2a)\quad (\Sigma_i \alpha_i N_{i,k})^{(j)} = \Sigma_i \alpha_i^{(j)} N_{i,k-j}$$

with

$$(5.2b)\quad \alpha_i^{(j)} := \begin{cases} \alpha_i, & j = 0 \\ \dfrac{\alpha_i^{(j-1)} - \alpha_{i-1}^{(j-1)}}{(t_{i+k-j} - t_i)/(k-j)}, & j > 0 \end{cases}$$

The recurrence relation (4.9) (with $j = 0$) allows one to

express a B-spline series as a series of lower order, but with polynomial coefficients. Precisely,

$$(5.3a)\qquad \Sigma_i \alpha_i N_{i,k}(x) = \Sigma_i \alpha_i^{[j]}(x) N_{i,k-j}(x)$$

with

$$(5.3b)\qquad \alpha_i^{[j]}(x) := \begin{cases} \alpha_i & , \ j = 0 \\ \dfrac{(x-t_i)\alpha_i^{[j-1]}(x) + (t_{i+k-j}-x)\alpha_{i-1}^{[j-1]}(x)}{t_{i+k-j} - t_i}, & j > 0. \end{cases}$$

In particular, $\alpha_i^{[k-1]}$ is a polynomial of degree $< k$ which agrees with $\Sigma_i \alpha_i N_{i,k}$ on $[t_i^+, t_{i+1}^-]$. Hence, (5.3) can be used to evaluate $\Sigma_i \alpha_i N_{i,k}$ at $x \in [t_i^+, t_{i+1}^-]$ by repeated formation of averages, starting with the k numbers $\alpha_{i-k+1}, \ldots, \alpha_i$ (see the first algorithm in [8]).

The quasi-interpolant of Fix and the author [16] provides an ofttimes convenient means for computing the B-spline coefficients of a given spline. The quasi-interpolant makes use of the linear functional λ_i given by the rule

$$(5.4)\qquad \lambda_i f := \lambda_{\tau_i, \psi_{i,k}} f := \sum_{j<k} (-)^{k-1-j} \psi_{i,k}^{(k-1-j)}(\tau_i) f^{(j)}(\tau_i).$$

Here,

$$\psi_{i,k}(x) := (t_{i+1}-x)\cdot \ldots \cdot (t_{i+k-1}-x)/(k-1)!$$

and τ_i is an <u>arbitrary</u> point in (t_i, t_{i+k}). Then, as one verifies directly [16],

$$(5.5)\qquad \lambda_i N_j = \delta_{i,j}, \quad \text{all } j.$$

Since λ_i has support at a point only, it follows that $\lambda_i(\Sigma_j \alpha_j N_j) = \alpha_i$.

The usefulness of this functional was demonstrated in [9]. For instance, it provides a quick proof of Theorem 3.1 and its

corollary. As another instance, it provides a quick proof of the fact due to Curry and Schoenberg [30] that B-splines are splines of minimal support: If $f \in \$_{k,\underline{\underline{t}}}$ has its support in (t_r, t_{r+s}) and $s < k$, then, for each i, one can choose τ_i in $(t_i, t_{i+k}) \setminus (t_r, t_{r+s})$, hence then $\lambda_i f = 0$, all i, i.e., $f = 0$.

More generally, one obtains

LEMMA 5.1. If $t_i < t_{i+k}$, all i, then $\overline{\text{supp}(\Sigma_i \alpha_i N_i)} = \bigcup_{\alpha_i \neq 0} \text{supp } N_i$.

In order to compute the coefficients of specific splines, we observe that, for $f, \psi \in \mathbb{P}_k$, $\alpha(\tau) := \lambda_{\tau,\psi} f$ is constant as a function of τ, as is clear from the fact that $\alpha'(\tau) = \psi(\tau) f^{(k)}(\tau) - (-)^k \psi^{(k)}(\tau) f(\tau)$. Hence, with $\tau = y$, we get that

$$\lambda_i (y-\cdot)^{k-1} = \lambda_{y,\psi_{i,k}} (y-\cdot)^{k-1} = \psi_{i,k}(y)(-)^{k-1}(k-1)! \;.$$

This shows that

$$(5.6) \quad (y-x)^{k-1} = \Sigma_i (y-t_{i+1}) \cdot \ldots \cdot (y-t_{i+k-1}) N_{i,k}(x),$$

which is Marsden's identity [61]. More generally,

$$\lambda_i (y-\cdot)^{k-p}/(k-p)! = (-)^{p-1} \psi_{i,k}^{(p-1)}(y)(-)^{k-p},$$

so

$$(5.7) \quad (y-x)^{k-p}/(k-p)! = (-)^{k-1} \Sigma_i \psi_{i,k}^{(p-1)}(y) N_{i,k}(x),$$

and, in particular, with $p = k$,

$$(5.8) \quad 1 = \Sigma_i N_{i,k}.$$

Of course, all these identities hold on $(t_{-\infty}, t_\infty)$ only. One obtains similarly that

(5.9) $\quad (y-x)_+^{k-1} = \Sigma_i (y-t_{i+1})_+ \cdot \ldots \cdot (y-t_{i+k-1})_+ N_{i,k}(x), \quad \text{for } y \in \underline{\underline{t}}.$

For the uniform knot sequence $\underline{\underline{t}} = \mathbb{Z}$ and for $k = 4$, one can find (5.6) and (5.9) already in Schoenberg [70].

Identities (5.6) and (5.9) illustrate a point to be made repeatedly in this survey, viz how closely a spline function is modelled by its B-spline coefficients. To elaborate on this point a little, note that, with $\underline{\underline{\tau}} := (\tau_i)_0^k$ any subsequence of $\underline{\underline{t}}$, (5.9) implies that

(5.10a) $\quad k[\tau_0,\ldots,\tau_k](\cdot - x)_+^{k-1} = \Sigma_i \alpha_{\underline{\underline{\tau}}}(i) M_{i,k}(x)$

where

(5.10b) $\quad \alpha_{\underline{\underline{\tau}}}(i) := (t_{i+k}-t_i)[\tau_0,\ldots,\tau_k](\cdot - t_{i+1})_+ \cdot \ldots \cdot (\cdot - t_{i+k-1})_+$
≥ 0

This supplies the formula

(5.11) $\quad [\tau_0,\ldots,\tau_k] = \Sigma_i \alpha_\tau(i)[t_i,\ldots,t_{i+k}]$

for the k-th divided difference at some points in terms of the k-th divided differences at the points of a refinement of those points, with the coefficients nonnegative. The existence of such a formula with nonnegative weights $\alpha_{\underline{\underline{\tau}}}$ was already known to Favard [35]. The formula is clearly a discrete analog of (4.1), and $\alpha_{\underline{\underline{\tau}}}$ deserves to be called a <u>discrete B-spline with knots</u> $\underline{\underline{\tau}}$. Indeed, $\alpha_{\underline{\underline{\tau}}}$ has been called just that by Schumaker [79] in the special case when $\underline{\underline{t}}$ is uniform, $t_i = t_0 + ih$, all i. In that case, if $f \in \mathcal{S}_{k,\underline{\underline{t}}}$ has only the active knots $t_{i_0},\ldots,t_{i_r}$ and $f = \Sigma_i \alpha_f(i) N_{i,k}$, then α_f is a discrete spline of order k with knots $i_0,\ldots,i_r$ in the sense of Mangasarian and Schumaker [60]. This means that, for each j, $\alpha_f(i)$ is a polynomial of order k in i on $i_j - k < i < i_{j+1}$. It should be said, though, that Mangasarian and Schumaker did

not view discrete splines in this light as B-spline coefficients of continuous splines. They arrived at discrete splines as the solution of certain discrete minimization problems.

The size of the i-th B-spline coefficient of a spline is closely tied (at least for moderate k) to the size of that spline "nearby," i.e., on (t_i, t_{i+k}), as can be proved [9] with the aid of the linear functional (5.4). Slightly more refined arguments produce the following explicit result.

THEOREM 5.1 [13]. Let D_k be the smallest number with the property that for every $\underline{\underline{t}}$, every i, and every $a < b$ with

$$t_i \le a \le t_{i+1}, \quad t_{i+k-1} \le b \le t_{i+k},$$

there exists $h_i \in \mathbb{L}_\infty$ such that

$$\text{(5.12)} \qquad \operatorname{supp} h_i \subseteq [a,b], \ \|h_i\|_\infty \le D_k/(b-a), \ \int h_i N_j = \delta_{ij}, \ \text{all } j.$$

Then $(\pi/2)^k/2 \le D_k \le 2k\, 9^{k-1}$.

Numerical evidence presented in [13] strongly suggests that actually $D_k \sim 2^k$.

The theorem implies that

$$\text{(5.13)} \qquad |\alpha_i|(t_{i+k}-t_i)^{1/p} \le D_k \|\Sigma_i \alpha_i N_i\|_{p,[t_i,t_{i+k}]}, \quad 1 \le p \le \infty,$$

which leads to

THEOREM 5.2 [9]. Let E be the diagonal matrix $\lceil \ldots, (t_{i+k}-t_i)/k, \ldots \rfloor$. Then

$$D_k^{-1} \|E^{1/p}\alpha\|_p \le \|\Sigma_i \alpha_i N_i\| \le \|E^{1/p}\alpha\|_p, \quad \text{all } \alpha \in \mathbb{R}^{\mathbb{Z}}, \ 1 \le p \le \infty .$$

In particular, $\Sigma_i \alpha_i N_i \in \mathbb{L}_p$ if and only if $E^{1/p}\alpha \in \ell_p(\mathbb{Z})$.

The proof of the upper bound for $\|\Sigma_i \alpha_i N_i\|_p$ makes use of the fact that the N_i's are nonnegative and sum up to 1 while, by (4.2) and (4.4), $\int N_{i,k} = (t_{i+k}-t_i)/k$.

COROLLARY 1 [9]. <u>Let</u> $N_{i,k,p} := (k/(t_{i+k}-t_i))^{1/p} N_{i,k}$. <u>For</u> $1 \le p < \infty$, $(N_{i,k,p})$ <u>is a Schauder basis for</u> $\mathcal{S}_{k,\underline{\underline{t}}} \cap \mathbb{L}_p(\mathbb{R})$.

We note the estimates

(5.14) $k^{1/p}/k \le \|N_{i,k,p}\|_p \le 1.$

COROLLARY 2 [12]. <u>Let</u> $\underline{\underline{t}}$ <u>be finite, infinite, or bi-infinite, let</u> $G := (\int N_{i,k,2} N_{j,k,2})$, <u>and let</u> $G^{-1} = (\alpha_{ij})$. <u>Then</u> G^{-1} <u>decays exponentially away from the diagonal. Explicitly,</u>

$$|\alpha_{ij}| \le \text{const } q^{|i-j|}$$

<u>with</u> $q = (1-D_k^{-2})^{1/(2k-2)} \in (0,1)$ <u>and</u> $\text{const} = D_k^3/q^{k-1}$ <u>both depending only on</u> k <u>and not on</u> $\underline{\underline{t}}$.

This corollary was proved earlier for a finite <u>uniform</u> $\underline{\underline{t}}$ by Domsta [34], and then used by Cielsielsky and Domsta [26] in the construction of a basis for $C^{k-2}[0,1]^d$ which is, at the same time, also a basis for $\mathbb{L}_p^{k-2}[0,1]^d$ for $1 \le p < \infty$. The corollary was used in [12] for a somewhat related purpose, viz in order to show that least-squares approximation from $\mathcal{S}_{k,\underline{\underline{t}}}$, considered as a map on $\mathbb{L}_p$, can be bounded in terms of the global mesh ratio

$$M_{\underline{\underline{t}}} := \sup_{i,j} (t_{i+k}-t_i)/(t_{j+k}-t_j).$$

COROLLARY 3 [7], [13]. <u>Let</u> $m\mathcal{S}_{k,\underline{\underline{t}}} := \mathcal{S}_{k,\underline{\underline{t}}} \cap m(\mathbb{R})$ <u>be the subspace of bounded splines of order</u> k <u>with knot sequence</u> $\underline{\underline{t}}$. <u>Then the rule</u> $\alpha \mapsto \Sigma_i \alpha_i N_i$ <u>maps</u> $\ell_\infty(\mathbb{Z})$ <u>onto</u> $m\mathcal{S}_{k,\underline{\underline{t}}}$. <u>Further, with</u> $\emptyset$: $\ell_\infty(\mathbb{Z}) \to m\mathcal{S}_{k,\underline{\underline{t}}}: \alpha \mapsto \Sigma_i \alpha_i N_i$, <u>the condition (number)</u> $\text{cond}_{k,\underline{\underline{t}}} := \|\emptyset\| \|\emptyset^{-1}\|$ <u>of the basis</u> (N_i) <u>for</u> $m\mathcal{S}_{k,\underline{\underline{t}}}$ <u>is bounded by</u> D_k <u>independent of</u> $\underline{\underline{t}}$.

Since $(D_1, D_2, D_3, D_4, \ldots) \leq (1, 2.5, 5.3, 10.1, \ldots)$, this shows the B-spline basis to be well conditioned, independent of $\underline{\underline{t}}$, for "small" k.

Finally, for another illustration of the fact that B-spline coefficients "model" the function they represent, observe that, for the particular choice

$$(5.15) \quad \tau_i = \tau_i^* := (t_{i+1} + \ldots + t_{i+k-1})/(k-1),$$

the coefficient of $f^{(1)}(\tau_i)$ in (5.4) vanishes. Then

$$\lambda_i f = f(\tau_i^*) + b_i$$

with

$$|b_i| = \left|\sum_{j=2}^{k-1} (-)^{k-1-j} \psi_{ik}^{(k-1-j)}(\tau_i^*) f^{(j)}(\tau_i^*)\right|$$

$$\leq \mathrm{const}_k \; (\max \Delta t_r)^2 \max_{2 \leq j < k} \|f^{(j)}\|_\infty .$$

Therefore, if, e.g., f is a fixed spline with $\|f^{(j)}\|_\infty < \infty$ for $2 \leq j < k$, and we write f as a linear combination of B-splines on a knot sequence $\underline{\underline{t}}$ which refines the knot sequence for f, then the resulting B-spline coefficient sequence α for f satisfies

$$\alpha_i = f(\tau_i^*) + O(\max \, (\Delta t_r)^2).$$

6. Local Spline Approximation

Because of their local support, B-splines have been instrumental in the construction of local spline interpolation and approximation schemes. In such a scheme, the approximation is taken in the form

$$(6.1) \quad Af := \Sigma_i (\mu_i f) N_i$$

with μ_i a linear functional with support in supp $N_i = (t_i, t_{i+k})$.

Since then $(Af)\big|_{(t_j,t_{j+1})}$ depends only on $f\big|_{(t_{j+1-k},t_{j+k})}$, such an approximation scheme is capable of reflecting, and taking advantage of, the local behavior of f.

LEMMA 6.1. <u>If</u> A <u>reproduces</u> $\mathbf{P}_k$ <u>on</u> $(t_{-\infty},t_\infty)$, <u>then</u>

$$(6.1)\qquad \|f - Af\|_{\infty,(t_j,t_{j+1})} \le (\sup_i \|\mu_i\|)\ \mathrm{dist}_{\infty,(t_{j+1-k},t_{j+k})}(f,\mathbf{P}_k).$$

The condition that A reproduce $\mathbf{P}_k$ is certainly satisfied in case A is a projector. This will happen iff (μ_i) is dual to (N_i), i.e., $\mu_i N_j = \delta_{ij}$, all i, j. In such a case, Af interpolates f at (μ_i) in the sense that $\mu_i Af = \mu_i f$, all i. A linear functional μ_i satisfying

$$(6.2)\qquad \mathrm{supp}\ \mu_i \subseteq \mathrm{supp}\ N_i = (t_i,t_{i+k}),\ \mu_i N_j = \delta_{ij},\ \text{all } j,$$

seems to have been constructed for the first time in [5], for the purpose of demonstrating the linear independence over an interval of all B-splines which do not vanish identically on that interval. Since then, such linear functionals have been constructed in various ways and for a variety of jobs. A summary and detailed discussion is given in [13].

The first local spline interpolation scheme seems to have been Birkhoff's local spline approximation by moments [4]. A corrected and extended version can be found in [6]. The scheme was not given in the form (6.1). It was therefore somewhat of a surprise to find that local spline approximation by moments is a special case of the quasi-interpolant of Fix and the author [16], i.e., of the form (6.1) with $\mu_i = \lambda_i$ given by (5.4) with $\tau_i = t_{i+k/2}$, all i.

The quasi-interpolant approximates well to f <u>and</u> its first k-1 derivatives, but requires values of f and of its derivatives for its construction. An earlier scheme [7]

constructs μ_i involving only function evaluations, and satisfying even

$$(6.3)\quad \operatorname{supp}\mu_i \subseteq (t_{i+1}, t_{i+k-1}), \quad \mu_i N_j = \delta_{ij}, \quad \text{all } j,$$

and so that $\sup_i \|\mu_i\| < \infty$. This is possible since it can be shown that

$$D_{k,\infty} := \sup_{\underline{\underline{t}}} \sup_i \inf\{\|\mu\| \mid \mu \in C^*[t_{i+1}, t_{i+k-1}],\ \mu N_j = \delta_{ij}, \text{ all } j\}$$

$$(6.4)\qquad = \sup_{\underline{\underline{t}}} \sup_i 1/\operatorname{dist}_{\infty,[t_{i+1},t_{i+k-1}]}(N_i, \operatorname{span}(N_j)_{j \neq i})$$

is finite. In fact, it follows from Theorem 5.1 that $D_{k,\infty} \leq D_k < \infty$. Therefore, one finds that, for this scheme,

$$(6.5)\quad \|f - Af\|_{\infty,(t_j,t_{j+1})} \leq D_{k,\infty} \operatorname{dist}_{\infty,(t_{j+2-k},t_{j+k-1})}(f, \mathbb{P}_k).$$

But it is not clear how well the derivatives of Af approximate those of f. Also, A is not applicable to arbitrary $f \in \mathbb{L}_p$.

The latter objection can be overcome by choosing μ_i of the form

$$\mu_i f = \int f h_i,$$

with $h_i \in \mathbb{L}_\infty[t_i, t_{i+k}]$ chosen as in Theorem 5.1 to satisfy (5.12). The resulting linear projector P,

$$(6.6)\quad Pf := \Sigma_i \left(\int f h_i\right) N_i,$$

is local and is bounded as a map on $\mathbb{L}_p$ by D_k for each $p \in [1,\infty]$ and independently of $\underline{\underline{t}}$ [11]. But, in order to obtain also good approximations to derivatives (regardless of $\underline{\underline{t}}$, i.e., without recourse to Markov's inequality), Lyche and Schumaker [59] found it necessary to give up the condition that Af interpolate f and to revert to the weaker condition that A merely reproduce $\mathbb{P}_k$. Such local approximation schemes have been further investigated by Demko [31].

An important local spline approximation scheme (which only reproduces $\mathbb{P}_2$) is Schoenberg's variation diminishing spline approximation. It will be discussed in the next section.

The use of local spline approximation schemes for gauging accurately the degree of approximation by splines is further pursued in DeVore's contribution to these proceedings.

We close this section with the remark that the dual to the linear projector P in (6.6), i.e., the linear projector P' given by

$$(6.7) \quad P'g := \Sigma_i (\int f N_i) h_i,$$

is helpful in settling two questions of "smooth" interpolation. The first, raised originally by Schoenberg [74] and partially answered by Golomb [42], concerns the existence of $g \in \mathbb{L}_p^k(\mathbb{R})$ which satisfies $g(t_i) = \alpha_i$, all i, for a given $\alpha \in \mathbb{R}^{\mathbb{Z}}$ and a given $\underline{\underline{t}} = (t_i)$ taken strictly increasing for simplicity. Let $[t_i,\ldots,t_{i+k}]\alpha$ be the k-th divided difference of the data at $t_i,\ldots,t_{i+k}$ and recall the diagonal matrix $E := \lceil\ldots, (t_{i+k}-t_i)/k,\ldots\rfloor$ of the preceding section. Then it is easily seen that having $E^{1/p}([t_i,\ldots,t_{i+k}])$ in ℓ_p is a necessary condition for the existence of such a g. To see that this is also a sufficient condition, observe [13] that the function g, given by the conditions that $g(t_i) = \alpha_i$, $i = 1,\ldots,k$ and that

$$(6.8) \quad g^{(k)} = (k-1)!\ \Sigma_i([t_i,\ldots,t_{i+k}]\alpha)(t_{i+k}-t_i)h_i,$$

is in $\mathbb{L}_p^k$ by Theorem 5.1 in case $E^{1/p}([t_i,\ldots,t_{i+k}]) \in \ell_p$, and agrees with α at $\underline{\underline{t}}$ since, by (4.2), it has the same k-th divided differences at the points of $\underline{\underline{t}}$ as does α.

The particular interpolant g to the given data α at $\underline{\underline{t}}$ just constructed has the property that, on $[t_j,t_{j+1}]$, at most k of the h_i in (6.8) are not zero, while, by Theorem 5.1, $\|h_i\|_\infty(t_{i+k}-t_i) \leq D_k$, all i. This proves [13] that, for given

$\underline{\underline{t}}$ and given α, there exists $g \in \mathbb{L}_\infty^k$ so that $g\big|_{\underline{\underline{t}}} = \alpha$ and, for all $t_j < t_{j+1}$,

$$\|g^{(k)}\|_{\infty,[t_j,t_{j+1}]} \le \text{const} \max_{[t_j,t_{j+1}]\subseteq[t_i,t_{i+k}]} k!\,|[t_i,\ldots,t_{i+k}]\alpha|$$

for some const $\le D_k$. This answers a question by H.-O. Kreiss as to the existence and the size of such a const.

7. Total Positivity and the Variation Diminishing Properties of B-splines

The strict positivity of $N_{i,k}$ on (t_i, t_{i+k}) (see (4.10)) is a particular instance of the Schoenberg-Whitney theorem and the variation diminishing properties of B-splines, the subject of this section. A thorough discussion of these matters in the more general context of Chebyshev splines can be found in Chapter 10 of Karlin's book on total positivity [47].

Throughout this section, the knot sequence is taken to be finite,

$$\underline{\underline{t}} = (t_i)_1^{n+k}, \text{ nondecreasing with } t_i < t_{i+k}, \text{ all } i,$$

and $(N_i)_1^n$ is the corresponding sequence of B-splines of order k. $\mathcal{S}_{k,\underline{\underline{t}}}$ has then dimension n. We consider spline interpolation at points $\tau_1 < \ldots < \tau_n$. This amounts to finding, for given f, $\alpha \in \mathbb{R}^n$ so that

$$(7.1) \qquad \sum_{j=1}^{n} \alpha_j N_j(\tau_i) = f(\tau_i), \quad i = 1,\ldots,n.$$

The question of existence and uniqueness of such an interpolant was settled some time ago.

THEOREM 7.1 (Schoenberg-Whitney [78]). Let

$$(7.2) \qquad S := \left\{ \sum_{j=1}^{k} \alpha_j x^{j-1} + \sum_{j=k+1}^{n} \alpha_j (x-t_j)_+^{k-1} \;\middle|\; \alpha \in \mathbb{R}^n \right\}$$

with $t_{k+1} < \ldots < t_n$. *If* $\tau_1 < \ldots < \tau_n$, *then* S *contains, for arbitrary* f, *an* s *such that* $s(\tau_i) = f(\tau_i)$, $i = 1,\ldots,n$ *iff* $\tau_{i-k} < t_i < \tau_i$, $i = k+1, \ldots, n$.

In this connection, it is interesting to note the following theorem published with an elegant proof in 1939, and pointed out to me by Allan Pinkus.

THEOREM (Krein and Finkelstein [55]). *Let* G *be a Green's function for the k-th order linear differential operator*

$$L = \sum_{j=0}^{k} p_j D^j$$

with $p_j \in C[a,b]$, *all* j, *and* p_k *never zero on* $[a,b]$. *Specifically, assume that* G *is of the form*

$$G(x,y) = \begin{cases} \sum_{j=1}^{p} \phi_j(x)\psi_j(y) & \text{for } x > y, \\ \sum_{j=1}^{q} \hat{\phi}_j(x)\hat{\psi}_j(y) & \text{for } x < y, \end{cases}$$

with both $(\phi_j)_1^p$ *and* $(\hat{\phi}_j)_1^q$ *linearly independent and in* ker L.

If $\det G\begin{pmatrix} x_1,\ldots,x_r \\ y_1,\ldots,y_r \end{pmatrix} \geq 0$ *for all nondecreasing* $(x_i)_1^r$ *and* $(y_i)_1^r$, *then*

$$\det G\begin{pmatrix} x_1,\ldots,x_r \\ y_1,\ldots,y_r \end{pmatrix} > 0 \quad \textit{for an increasing} \quad (x_i)_1^r,\ (y_i)_1^r$$

if and only if $x_{i-p} < y_i$, $i = p+1,\ldots,r$, *and* $y_i < x_{i+q}$, $i = 1,\ldots,r-q$.

Since S, as defined in (7.2), agrees with $\mathcal{S}_{k,\underline{\underline{t}}}$ on $[t_k, t_{n+1}]$, it is possible to translate Theorem 7.1 into a statement involving B-splines provided we make the assumption that

(7.3) $\tau_1, \ldots, \tau_n \in [t_k, t_{n+1}]$.

It is also possible to prove directly

THEOREM 7.2. _If_ $\tau_1 < \ldots < \tau_n$, _then_ $(N_j(\tau_i))_1^n$ _is invertible if and only if_ $\tau_i \in \operatorname{supp} N_i$, _i.e._, $N_i(\tau_i) \neq 0$, _all_ i.

In other words, $(N_j(\tau_i))$ is invertible iff its diagonal is invertible. Burchard [21, Chap. III, 2(3)] and Karlin [47, Chap. 10, Lemma 4.1] both prove Theorem 7.2 explicitly in terms of B-splines, with simple knots, but, on the other hand, more generally for Chebyshev splines.

Karlin and Ziegler [53] remove the restriction in Theorem 7.2 to simple knots. They also allow for repeated or osculatory interpolation and consider Chebyshev splines rather than just polynomial splines. Straightforward translation of their result to B-splines would require assumption (7.3).

We will now quit belaboring this minor point and state the theorem directly in terms of B-splines.

THEOREM 7.3 (Karlin-Ziegler [53] extension of Schoenberg-Whitney). _Let_ $\tau_1 \leq \ldots \leq \tau_n$ _be such that_

$$\tau_{i+1} = \ldots = \tau_{i+r} = t_{j+1} = \ldots = t_{j+s} \quad \underline{\text{implies}} \quad r + s \leq k,$$

and define linear functionals $(\mu_i)_1^n$ _by the rule_

$$\mu_i f := f^{(j)}(\tau_i) \quad \underline{\text{with}} \quad j := \max\{r \mid \tau_{i-r} = \tau_i\}.$$

Then $(\mu_i N_j)$ _is invertible if and only if_ $N_i(\tau_i) \neq 0$, $i=1,\ldots,n$.

A simple proof of this theorem, using only elementary properties of B-splines and Rolle's theorem, can be found in [15].

Theorem 7.3 states conditions under which it is possible to interpolate by linear combinations of _all_ B-splines for a given knot sequence. A careful study of Karlin's proof [47] of the total positivity of $(N_j(\tau_i))$ reveals the fact that

Theorem 7.3 remains valid if we replace the sequence (N_j) by one of its subsequences.

THEOREM 7.4 [15]. Under the same assumptions as those of Theorem 7.3, and for any subsequence $(q_1,\ldots,q_m)$ of $(1,\ldots,n)$, $\det(\mu_i N_{q_j})_{i,j=1}^m \geq 0$ with equality iff, for some i, $N_{q_i}(\tau_i) = 0$.

This theorem implies at once the total positivity of $(N_j(\tau_i))$.

THEOREM 7.5 (Karlin [47]). Let $\tau_1 \leq \ldots \leq \tau_n$. Then $(N_j(\tau_i))$ is totally positive, i.e., all its minors are nonnegative.

Karlin [47, p. 563] states that this theorem was communicated to him by Schoenberg.

COROLLARY. (N_i) is a weak Descartes system, i.e., any subsequence $(N_{q_i})_1^m$ of $(N_i)_1^n$ is a weak Chebyshev system.

The total positivity of $(N_j(\tau_i))$ provides bounds on the effect of rounding errors when solving (7.1) by Gauss elimination without pivoting which are smaller than those obtainable for general matrices even when using pivoting [18]. This means that it is reasonable to solve the banded system (7.1) without pivoting with the attendant savings in storage and program complexity.

The total positivity of $(N_j(\tau_i))$ is used in an essential way by Karlin and Pinkus [51] in their extension to splines and to higher derivatives of earlier results by C. Davis and Videnski concerning the existence of a polynomial of degree n on [0,1] with a prescribed sequence of n+1 extrema.

The total positivity of $(N_j(\tau_i))$ leads to one of the more striking spline approximation schemes, Schoenberg's

variation diminishing spline approximation, which has found much use in computer-aided design (see, e.g., Riesenfeld [69]). We recall some notation. A real-valued function f on some subset D of $\mathbb{R}$ has at least m strong sign changes if f alternates (in sign) on some $(\tau_i)_0^m$ in D, i.e., if

$$f(\tau_0) \neq 0 \text{ and, in case } m > 0,\ f(\tau_{i-1})f(\tau_i) < 0 \text{ for } i=1,\ldots,m,$$

for some nondecreasing sequence $(\tau_i)_0^m$ in D. It is customary to denote by

$$S^-(f)$$

the total number of strong sign changes of f on its domain. It is well known (e.g., Theorem 5.1.4 of [47]) that, for a totally positive matrix A and any vector α,

$$S^-(A\alpha) \leq S^-(\alpha),$$

i.e., a totally positive matrix transformation is variation diminishing. Since $(N_j(\tau_i))$ is totally positive, it follows that the linear map V_τ, given for some nondecreasing τ by

$$(7.4)\quad V_\tau f := \sum_{j=1}^{n} f(\tau_j)N_j, \quad \text{all } f,$$

is variation diminishing, i.e., $S^-(V_\tau f) \leq S^-(f)$. Recall now from Marsden's identity (see (5.6) and (5.7)) that, for any straight line p and any τ with $\tau_i \in (t_i, t_{i+k})$, all i,

$$p = \sum_{j=1}^{n} \left\{ p(\tau_j) + p'(\tau_j)\left[\sum_{r=1}^{k-1} t_{j+r} - (k-1)\tau_j\right]/(k-1)\right\} N_j$$

on $[t_k, t_{n+1}]$. Therefore, with the particular choice

$$(7.5)\quad \tau_j^* := (t_{j+1} + \ldots + t_{j+k-1})/(k-1), \quad j = 1,\ldots,n,$$

mentioned already in (5.15), V_{τ^*} reproduces $\mathbb{P}_2$ on $[t_k, t_{n+1}]$, and we have

(7.6) $\quad S^-(V_{\tau^*}f-p) \le S^-(f-p)$ on $[t_k, t_{n+1}]$, all $p \in \mathbb{P}_2$, all f.

The resulting approximation $V_{\tau^*}f$ to f is <u>Schoenberg's variation diminishing spline approximation</u>, introduced by Schoenberg in [73] and further discussed in Marsden and Schoenberg [63].

We note the following result due to Marsden [62]: Write $V_{\tau^*,k}$ to stress dependence on k, and restrict $\underline{\underline{t}}$ so that $t_i = \ldots = t_k = 0$ and $t_{n+1} = \ldots = t_{n+k} = 1$. Then

(7.7) $\quad V_{\tau^*,k} \to 1$ <u>pointwise on</u> $C[0,1]$ <u>iff</u> $\max_i \Delta t_i/k \to 0$,

as Marsden shows with the aid of the Bohman-Korovkin theorem concerning strong convergence of positive operators to the identity on $C[0,1]$.

It is possible to refine the proof that $S^-(A\alpha) \le S^-(\alpha)$ for a totally positive matrix A for the particular choice $A = (N_j(\tau_i))$ so as to obtain the following theorem.

THEOREM 7.6 [15]. <u>If</u> $f := \sum_{j=1}^{n} \alpha_j N_j$ <u>alternates on</u> $(\tau_i)_0^m$, <u>then</u>

$$f(\tau_i)\alpha_{q_i} N_{q_i}(\tau_i) > 0, \quad i = 0,\ldots,m,$$

<u>for some subsequence</u> q <u>of</u> $(1,\ldots,n)$.

Theorem 7.6 illustrates the point made earlier that B-spline coefficients "model" the function they represent. A spline cannot change sign at a point without its B-spline sequence also changing sign "nearby."

As a specific application of this theorem, consider the spline $N_i^{(j)}$ which, by (5.2), is the linear combination of $j+1$ B-splines (of order $k-j$), hence cannot have more than j strong sign changes, by Theorem 7.6. On the other hand, if $N_i^{(j-1)}$ is continuous, hence absolutely continuous, then $N_i^{(j)}$ is orthogonal to $\mathbb{P}_j$ on $[t_i, t_{i+k}]$, by (4.15), therefore must have at least j strong sign changes.

COROLLARY [30]. <u>B-splines are bell-shaped. Precisely, if</u> $N_i^{(j-1)}$ <u>is continuous for some</u> $j < k$, <u>then</u> $N_i^{(j)}$ <u>has exactly</u> j <u>zeros in</u> (t_i, t_{i+k}), <u>all simple, i.e., there exists</u> $(\xi_r)_0^{j+1}$ <u>with</u> $t_i = \xi_0 < \ldots < \xi_{j+1} = t_{i+k}$ <u>so that</u> $(-)^r N_i^{(j)} > 0$ <u>on</u> (ξ_r, ξ_{r+1}), $r = 0,\ldots,j$.

Finally, we record the relationship between B-splines and Pólya frequency functions discovered by Curry and Schoenberg [30]. By definition, a Pólya frequency distribution is any distribution function F (i.e., any function of the form $F(x) = \int_{-\infty}^{x} f(s)ds$ with f nonnegative and $F(\infty) = 1$) whose bilateral Laplace transform is of the form

$$\int_{-\infty}^{\infty} e^{-sx} dF(x) = 1/\psi(s)$$

with

$$\psi(s) = e^{-\gamma s^2 + \delta s} \prod_1^{\infty} (1+\delta_v s) e^{-\delta_v s}$$

for some $\gamma \geq 0$, δ real, and $(\delta_v) \in \ell_2$. If $\psi(s) = e^{\delta s}$, then dF has its entire unit mass located at $x = \delta$. If $\psi(s) \neq e^{\delta s}$, then

$$\int_{-\infty}^{\infty} e^{-sx} \Lambda(x)\, dx = 1/\psi(s)$$

with Λ a Pólya frequency function, i.e., a nonnegative integrable function on $\mathbb{R}$ (normalized to have $\int \Lambda = 1$) for which the kernel

$$K(x,y) := \Lambda(x-y)$$

is totally positive of all orders.

Call F_k a spline distribution function of order k if F_k has a B-spline of order k as its density, i.e., if

$$F_k(x) = k \int_{-\infty}^{x} [\tau_0, \ldots, \tau_k](\cdot - s)_+^{k-1} ds$$

for some $\tau_0 \leq \ldots \leq \tau_k$ with $\tau_0 < \tau_k$. Note that $F_k(x) = 0$ for $x \leq \tau_0$ and $F_k(x) = 1$ for $x \geq \tau_k$, by (4.2). Further, say that F_k converges to a distribution function F in case $\lim_{k\to\infty} F_k(x) = F(x)$ for all points x at which F is continuous.

THEOREM 7.6 [30]. *The distribution function* F *is a Pólya frequency distribution iff* F *is the limit of a sequence* (F_k) *of spline distributions, with* F_k *of order* k, *all* k.

8. "Best" Interpolation

In this section, I finally discuss an aspect of splines which many consider to be the primary characteristic of splines, viz. the fact that splines are solutions to interesting variational problems. This property of splines is closely related to the fact that the B-spline $M_{i,k}$ represents a k-th order divided difference. As mentioned already in (4.1), if $a \leq t_i < t_{i+k} \leq b$, then

$$(8.1) \quad [t_i, \ldots, t_{i+k}]f = \int_a^b M_{i,k}(s)\ f^{(k)}(ds)/k!$$

for every $f \in \mathbb{M}^k[a,b] := \{f \in C^{k-2}[a,b] \mid f^{(k-2)}$ abs.cont., $f^{(k-1)} \in BV\}$.

Details for the material in this section can be found in [14] and its references.

Consider the problem of minimizing $\|f^{(k)}\|_p$ over

$$(8.2) \quad F_p := F_p(\underline{\underline{\tau}}, \alpha, k, [a,b]) := \{f \in \mathbb{L}_p^k[a,b] : f|_{\underline{\underline{\tau}}} = \alpha\}$$

for given $\underline{\underline{\tau}} := (\tau_i)_1^n$ in $[a,b]$, nondecreasing with $\tau_i < t_{i+k}$, all i, and given $\alpha \in \mathbb{R}^n$, with $[a,b]$ finite, positive $k \leq n$ and $p \in [1,\infty]$. Here, $f|_{\underline{\underline{\tau}}}$ is the sequence $(f_i)_1^n$ given by the rule

$$f_1 := f^{(j)}(\tau_i), \quad \text{with } j := \max\{ r \mid \tau_{i-r} = \tau_i \}.$$

F_p is not empty. It contains, e.g., exactly one polynomial of degree $< n$. Therefore,

$$F_p = \{f \in \mathbb{L}_p^k[a,b] \mid f|_{\underline{\underline{\tau}}} = f_\alpha|_{\underline{\underline{\tau}}}\}$$

for some fixed $f_\alpha \in F_p$. Favard [35] already knew and used the fact that

$$\inf_{f \in F_p} \|f^{(k)}\|_p = \inf_{g \in G_p} \|g\|_p$$

with

$$(8.2') \quad G_p := \{f^{(k)} \mid f \in F_p\} = \{g \in \mathbb{L}_p \mid \int M_{i,k} g = k![\tau_i, \ldots, \tau_{i+k}]f_\alpha, \quad i = 1,\ldots,n-k\}.$$

Let now $1 < p \leq \infty$ and $1/p + 1/q = 1$. Then, following Krein [54], we recognize that minimization of $\|g_p\|$ over G_p can be viewed, dually, as the construction of an extension $\lambda \in \mathbb{L}_p = \mathbb{L}_q^*$ of minimal norm to all of $\mathbb{L}_q[a,b]$ of the linear functional λ_α, given on $\mathcal{S}_{k,\underline{\underline{\tau}}} = \mathrm{span}(M_{i,k})_1^{n-k} \subseteq \mathbb{L}_q[a,b]$ by

$$\lambda_\alpha : \mathcal{S}_{k,\underline{\underline{\tau}}} \to \mathbb{R} : \Sigma_i \beta_i M_{i,k} \mapsto \Sigma_i \beta_i k![\tau_i, \ldots, \tau_{i+k}]f_\alpha .$$

This is so since G_p, as a subset of $\mathbb{L}_q^*$, coincides with the set of all extensions of λ_α. Therefore

$$(8.3) \quad \inf_{f \in F_p} \|f^{(k)}\|_p = \min\{\|\lambda\| \mid \lambda \in \mathbb{L}_q^*, \ \lambda|_{\mathcal{S}_{k,\underline{\underline{\tau}}}} = \lambda_\alpha\} = \|\lambda_\alpha\| ,$$

by the Hahn-Banach theorem, settling existence of a minimal f in F_p as well. Further, a minimal f must agree with f_α at $\tau_1, \ldots, \tau_k$ while its k-th derivative satisfies

$$(8.4) \quad \int_a^b f^{(k)}(s)\psi(s)\,ds = \|f^{(k)}\|_p \|\psi\|_q$$

for any $\mathbb{L}_q$-extremal ψ of λ_α, i.e., for any ψ with

$$(8.5) \quad \psi \in \mathcal{S}_{k,\underline{\underline{\tau}}} \quad \text{and} \quad \|\psi\|_q = 1 \quad \text{and} \quad \lambda_\alpha \psi = \|\lambda_\alpha\|.$$

If $\lambda_\alpha = 0$, then there is a polynomial of order k in F_p and it is the unique minimizer for all p. Otherwise $\lambda_\alpha \neq 0$. But then, for $1 < q < \infty$, λ_α has exactly one extremal and the equality (8.4) in Hölder's inequality then forces $f^{(k)}$ to satisfy

$$f^{(k)} = \|\lambda_\alpha\| \, |\psi|^{q-1} \operatorname{signum} \psi. \tag{8.6}$$

It follows that $\|f^{(k)}\|_p$ is uniquely minimized on F_p, and the minimizer is the unique element $\hat{f}_p$ of the nonlinear family

$$\{f \in \mathbb{L}_1^k[a,b] \mid f^{(k)} = |\psi|^{q-1} \operatorname{signum} \psi \text{ for some } \psi \in \mathcal{S}_{k,\underline{\underline{\tau}}}\}, \tag{8.7}$$

for which $\hat{f}_p\big|_{\underline{\underline{\tau}}} = f_\alpha\big|_{\underline{\underline{\tau}}}$. Such functions have been called $\mathbb{L}_p$-<u>splines</u> by Golomb [42] who was apparently the first to describe their structure.

For $p = 2$, the family (8.7) is linear and consists of all $f \in \mathbb{L}_1^k$ with $f^{(k)} \in \mathcal{S}_{k,\underline{\underline{\tau}}}$. To describe the corresponding minimizer, let $\underline{\underline{t}}$ be the extension of $\underline{\underline{\tau}}$ to a nondecreasing sequence having both a and b occurring exactly 2k times. Then the minimizer in F_2 is the unique $\hat{f}_2$ in $\mathcal{S}_{2k,\underline{\underline{t}}}$ which, in addition to the condition $\hat{f}_2\big|_{\underline{\underline{\tau}}} = f_\alpha\big|_{\underline{\underline{\tau}}}$, also satisfies

$$(\tau_i - a)\hat{f}_2^{(2k-i)}(a^+) = (b - \tau_{n+1-i})\hat{f}_2^{(2k-i)}(b^-) = 0, \quad i = 1,\ldots,k. \tag{8.8}$$

The minimizer has been called by Schoenberg [73] the <u>natural</u> spline interpolant, of order 2k with interior knots $\tau_1,\ldots,\tau_n$ for f_α in case $a < \tau_1$ and $\tau_n < b$, in which case all the constraints (8.8) on $\hat{f}_2$ are active. At the other extreme, when none of the constraints (8.8) on f_2 is active, i.e., when $a = \tau_1 = \ldots = \tau_k$ and $\tau_{n-k+1} = \ldots = \tau_n = b$, the minimizer has been called by Schoenberg (see, e.g., Lecture 7 of [76]) the <u>complete</u> spline interpolant for f_α of order $2k$ with interior knots $\tau_{k+1},\ldots,\tau_{n-k}$. The word "spline" itself

was chosen by Schoenberg [70] because in the case $k = 2$ the resulting interpolating cubic spline approximates (for small slopes) the position of a mechanical or draftman's spline forced to go through the given data points. This connection between (2k-1)st degree spline interpolation at knots and least-squares approximation to the k-th derivative has remained for many the major reason for using splines.

For $p = \infty$, (8.4) fails to pin down the minimizer uniquely since it only implies that

$$(8.9) \quad f^{(k)} = \|\lambda_\alpha\| \operatorname{signum} \psi \text{ off } N_\psi := \{x \in [a,b] \mid \psi(x) = 0\}$$

for every $\mathbb{L}_1$-extremal ψ of λ_α. Of course, if N_ψ has measure zero, then it follows that the minimizer $\hat{f}$ is unique and its k-th derivative is absolutely constant, with $< n-k$ break points, by Theorem 7.5, since ψ is a nontrivial linear combination of n-k B-splines. In the language of Glaeser [40,41], $\hat{f}$ is a <u>perfect spline of degree</u> k, i.e., a pp function of order $k+1$ in C^{k-1} with absolutely constant k-th derivative.

Whether or not N_ψ has zero measure, $\operatorname{supp} \psi = [a,b] \backslash N_\psi$ must contain the support of some B-spline of order k for the knot sequence $\underline{\underline{\tau}}$, by Lemma 5.1, i.e., some interval (τ_i, τ_{i+k}) on which then, by (8.9), all minimizers must agree. This is the "core interval of uniqueness" of Fisher and Jerome [36]. In particular, the minimizer is uniquely determined in case $n = k+1$. It is also uniquely determined in case $n = 2k$ and

$$a = \tau_1 = \ldots = \tau_k, \quad \tau_{k+1} = \ldots = \tau_{2k} = b,$$

as was found by Glaeser [40,41], since now $\mathcal{S}_{k,\underline{\underline{\tau}}} = \mathbb{P}_k \big|_{[a,b]}$. For the specific data $f_\alpha(x) := \int_a^x (s-a)^{k-1}(b-s)^{k-1} ds$, Louboutin [58] (see also Schoenberg [75,76]) found $\hat{f}$ explicitly in this case: $f_\alpha^{(k)}$ is evidently orthogonal to $\mathbb{P}_{k-1} \subseteq \mathcal{S}_{k,\underline{\underline{\tau}}}$ on $[a,b]$, therefore $\hat{f}^{(k)}$ must be a step function with $< k$ jumps and

orthogonal to $\mathbb{P}_{k-1}$ on $[a,b]$. But, since $\mathbb{P}_{k-1}$ is a Chebyshev system, this pins down signum $\hat{f}^{(k)}$ uniquely up to multiplication by a sign $\sigma \in \{-1,1\}$,

$$\text{signum } \hat{f}^{(k)} = \sigma \text{ signum } C_k^{(1)}$$

with $C_k(x) = (-)^{k-1} C_k(2\frac{x-a}{b-a} - 1)$ and C_k the Chebyshev polynomial of degree k. It follows that $f^{(1)}$ is a B-spline of order k with simple knots at the $k+1$ extrema of C_k on $[a,b]$ (see (4.15)). But, in general, there will be several distinct minimizers. Karlin [48] was the first to see that among these has to be at least one perfect spline $\hat{f}$ of degree k with $< n-k$ interior knots. Its derivative $\hat{f}^{(k)}$ can be constructed [10] as a limit point of the net $(g_\varepsilon)_{\varepsilon > 0}$, with g_ε the unique minimizer of $\| \, \|_\infty$ in

$$G_{\infty,\varepsilon} := \{g \in \mathbb{L}_\infty[a,b] \mid \int_a^b \phi g = \int_a^b \phi f_\alpha^{(k)}, \quad \text{all } \phi \in S_\varepsilon\}$$

where

$$S_\varepsilon := K_\varepsilon(\mathcal{S}_{k,\underline{\tau}}), (K_\varepsilon\phi)(x) := \int_{-\infty}^{\infty} \exp(-(y-x)^2/(2\varepsilon^2)\phi(y)\,dy/(\varepsilon\sqrt{2\pi}).$$

The minimizer g_ε is in fact uniquely determined, absolutely constant and has $< n-k$ jumps, since the total positivity of $(N_{j,k}(\sigma_i))$ for increasing $\underline{\sigma}$ (see Theorem 7.5) implies [47] that $(K_\varepsilon N_{j,k}(\sigma_i))$ is strictly totally positive for strictly increasing $\underline{\sigma}$; therefore any nonzero element ψ of S_ε vanishes on $< n-k$ points. Finally, Favard [35] constructed a minimizer $\hat{f}$ which is a spline of degree k with $< n-k$ interior knots, all simple, with the additional property that, for any $f \in F_\infty$, $|f^{(k)}| \le |\hat{f}^{(k)}|$ implies that $f = \hat{f}$. This minimality of "Favard's solution" is further underlined by the fact that it is, for any $r \in [1,\infty)$, the $\mathbb{L}_r^k$-limit of $\hat{f}_p$ as $p \to \infty$ [25].

For $p = 1$, matters are least satisfactory since $\mathbb{L}_p$ now fails to be the dual for $\mathbb{L}_q$. Therefore, although (8.3) still holds for this case, it may happen that none of the norm

preserving extensions of λ_α to all of $\mathbb{L}_\infty$ is representable as integration against an $\mathbb{L}_1$-function, in which case the infimum over F_1 is not attained. In this situation, one may be satisfied to follow the lead of Fisher and Jerome [37] and consider the slightly different problem of minimizing

$$\|f^{(k)}\| := \operatorname{Var} f^{(k-1)}$$

over

$$F_1 := \{f \in \mathbb{M}^k[a,b] \mid f|_{\underline{\underline{\tau}}} = f_\alpha|_{\underline{\underline{\tau}}}\}$$

instead, which always has solutions. If $\tau_i < \tau_{i+k-1}$, all i, then among these solutions is a spline of order k with $\leq n-k$ interior knots, all simple.

We close this section with yet another B-spline property, this one connected with perfect splines, optimal recovery (alias best class estimators) and $\mathbb{L}_1$-approximation by splines.

LEMMA 8.1 (Micchelli [64]). <u>If</u> $\underline{\underline{\tau}} = (\tau_i)_1^n$ <u>is nondecreasing in</u> (a,b) <u>with</u> $n > k$, <u>then there exists (up to multiplication by</u> <u>some</u> $\sigma \in \{-1,1\}$) <u>exactly one sign function</u> h <u>with</u> $\leq n-k$ <u>jumps which is orthogonal to</u> $\mathcal{S}_{k,\underline{\underline{\tau}}}$ <u>on</u> $[a,b]$. <u>If</u> $a = \xi_0 < \dots < \xi_{r+1} = b$, <u>and, for this</u> h, $(-)^i h = 1$ <u>on</u> (ξ_i, ξ_{i+1}), $i = 0,\dots,r$, <u>then</u> $r = n-k$ <u>and</u> $\xi_i \in (\tau_i, \tau_{i+k})$, $i = 1,\dots,r$.

Micchelli's lemma is not entirely unrelated to the following fact about B-splines useful, e.g., in the characterization of best $\mathbb{L}_p$-approximations by splines.

LEMMA 8.2. <u>If</u> $\underline{\underline{t}} = (t_i)_1^{n+k}$ <u>is nondecreasing, in</u> $[a,b]$, <u>with</u> $t_i < t_{i+k}$, <u>all</u> i, <u>and</u> $f \in \mathbb{L}_1[a,b]$ <u>is orthogonal to</u> $\mathcal{S}_{k,\underline{\underline{t}}}$ <u>on</u> $[a,b]$, <u>then there exists</u> $\underline{\underline{\xi}} = (\xi_i)_1^{n+1}$ <u>strictly increasing in</u> $[a,b]$ <u>with</u> $t_i \leq \xi_i \leq t_{i+k-1}$ <u>(any equality holding iff</u> $t_i = t_{i+k-1}$), $i = 1,\dots,n+1$, <u>so that</u> f <u>is also orthogonal to</u> $\mathcal{S}_{1,\underline{\underline{\xi}}}$.

Indeed, since, for appropriately chosen $p \in \mathbb{P}_k$, the

function $F := p + \int_a^b (\cdot - y)_+^{k-1} f(y)\,dy/(k-1)!$ vanishes at $\underline{\underline{t}}$ (counting multiplicities) by assumption, and F is in $C^{k-1}[a,b]$, Rolle's Theorem proves the existence of strictly increasing $(\xi_i)_1^{n+1}$ in $[a,b]$ with $t_i \le \xi_i \le t_{i+k-1}$, all i, at which $F^{(k-1)} = \text{const} + \int_a^b (\cdot - y)_+^0 f(y)\,dy$ vanishes, which proves the lemma. In particular, if f is continuous, then it must vanish at the n points of some strictly increasing sequence $(\eta_i)_1^n$ with $t_i < \eta_i < t_{i+k}$, all i.

9. Generalizations

The trend started by Schoenberg [71] and Greville [43] toward ever more generalized splines continues unabated but has failed to bring with it a corresponding wealth of generalized B-splines. Schoenberg [71] actually described trigonometric B-splines and later, Burchard [21] and Karlin [47] independently constructed Chebyshevian B-splines with the aid of Popoviciu's [67] generalization of the divided difference notion. Yet another account can be found in Marsden's thesis, eventually published in [61], in which the generalization of Schoenberg's variation diminishing spline approximation for Chebyshev splines is given, but without a proof of its variation diminishing character. Such a scheme had already been described and proven to be variation diminishing by Karlin and Karon [49], and their assertion in [50] that Marsden's B-splines are essentially different from Karlin's is incorrect.

Here are some of the details of the construction.

Let Pf be the polynomial of degree $< k$ which agrees with f at the distinct points $\tau_1,\ldots,\tau_k$. If $\phi_j(x) = x^{j-1}$, all j, then

$$(9.1)\quad f - Pf = \det\begin{pmatrix}\tau_1,\ldots,\tau_k, & \cdot\\ \phi_1,\ldots,\phi_k, & f\end{pmatrix} \Big/ \det\begin{pmatrix}\tau_1,\ldots,\tau_k\\ \phi_1,\ldots,\phi_k\end{pmatrix}.$$

Therefore, since $[\tau_1,\ldots,\tau_k,x]f$ is the leading coefficient in the polynomial of degree $\leq k$ which agrees with f at $\tau_1,\ldots,\tau_k,x$, we have

$$(9.2)\quad f - Pf = ([\tau_1,\ldots,\tau_k,\cdot]f)(\phi_{k+1}-P\phi_{k+1})$$

with

$$[\tau_1,\ldots,\tau_k,\cdot]f = (f-Pf)/(\phi_{k+1}-P\phi_{k+1})$$
$$= \det\begin{pmatrix}\tau_1,\ldots,\tau_k, & \bullet\\ \phi_1,\ldots,\phi_k, & f\end{pmatrix}\Big/\det\begin{pmatrix}\tau_1,\ldots,\tau_k, & \bullet\\ \phi_1,\ldots,\phi_{k+1}\end{pmatrix}.$$

If now, more generally, $(\phi_j)_1^{k+1}$ is a Chebyshev system (on some interval I), then $\det\begin{pmatrix}\tau_1,\ldots,\tau_{k+1}\\ \phi_1,\ldots,\phi_{k+1}\end{pmatrix} \neq 0$ for distinct $\tau_1,\ldots,\tau_{k+1}$ in I and the following definition makes sense: <u>The k-th divided difference of</u> f <u>at the distinct points</u> $\tau_1,\ldots,\tau_{k+1}$ <u>in</u> I <u>with respect to the sequence</u> $\phi := (\phi_j)_1^{k+1}$ <u>is</u> [67]

$$(9.3)\quad [\tau_1,\ldots,\tau_{k+1}]_\phi f := \det\begin{pmatrix}\tau_1,\ldots,\tau_{k+1}\\ \phi_1,\ldots,\phi_k,f\end{pmatrix}\Big/\det\begin{pmatrix}\tau_1,\ldots,\tau_{k+1}\\ \phi_1,\ldots,\phi_{k+1}\end{pmatrix}.$$

Then, with Pf denoting, more generally, the unique element in $\operatorname{span}(\phi_j)_1^k$ which agrees with f at $\tau_1,\ldots,\tau_{k+1}$, we have

$$f - Pf = ([\tau_1,\ldots,\tau_k,\cdot]_\phi f)(\phi_{k+1} - P\phi_{k+1})$$

which is the formal analog of (9.2). The definition shows the generalized divided difference (9.3) to be a symmetric function of the τ_i's. The definition even allows for some confluence among the τ_i's provided the ϕ_j's are sufficiently smooth and one defines (for nondecreasing $\underline{\underline{\tau}}$)

$$\det\begin{pmatrix}\tau_1,\ldots,\tau_{k+1}\\ \phi_1,\ldots,\phi_{k+1}\end{pmatrix} := \det\begin{pmatrix}\mu_1,\ldots,\mu_{k+1}\\ \phi_1,\ldots,\phi_{k+1}\end{pmatrix} = \det(\mu_i\phi_j)$$

with $\mu_i f := f^{(j)}(\tau_i)$ and $j := \max\{r \mid \tau_{i-r} = \tau_i\}$, in the manner

of Theorem 7.3. More detail about these generalized divided differences are provided by Popoviciu [67], and see also Mühlbach [65].

Assume that, in addition, $(\phi_j)_1^k$ spans the kernel of a k-th order linear ordinary differential operator

(9.4) $\quad L^* := D^k + \sum_{j<k} a_j D^j$

with $a_j \in C^j(I)$, all j, so that the formal adjoint

(9.5) $\quad L := (-)^k D^k + \sum_{j<k} (-)^j D^j (a_j \cdot) = (-)^k (D^k + \sum_{j<k} b_j D^j)$

is an operator of the same kind. Green's function $G(x,y)$ for the initial value problem $L^* f = g, f^{(j)}(a) = 0$, $j = 0,\ldots,k-1$, can then be constructed as

(9.6a) $\quad G(x,y) = (x-y)_+^0 \sum_{j=1}^k \phi_j(x)\psi_j(y)$

with $(\psi_j)_1^k$ the basis for ker L adjunct to $(\phi_i)_1^k$, i.e.,

(9.6b) $\quad \sum_{j=1}^k \phi_j^{(i-1)}(x)\psi_j(x) = \delta_{ik}, \quad i = 1,\ldots,k,\ x \in I.$

With $\underline{\underline{t}} = (t_i)_1^{n+k}$ nondecreasing and $t_i < t_{i+k}$, all i, the function

(9.7) $\quad M_{i,L}(y) := [t_i,\ldots,t_{i+k}]_\phi G(\cdot,y)$

is then piecewise in ker L with breakpoints $t_i,\ldots,t_{i+k}$, and in C^{k-2} in case $t_i < \ldots < t_{i+k}$. In the language of Greville [43], $M_{i,L}$ is a generalized spline function with respect to ker L. Coincidences among the t_i's reduce the smoothness of $M_{i,L}$ across t_j in the usual way. Further,

(9.8) $\quad M_{i,L}$ <u>vanishes off</u> (t_i, t_{i+k})

since, for $y > t_{i+k}$, $G(\cdot,y)\big|_{(t_i,t_{i+k})} = 0$ while, for $y < t_i$, $G(\cdot,y)\big|_{(t_i,t_{i+k})} \in \ker L^*$ by (9.6). One also has the analog

$$(9.9)\quad [t_i,\ldots,t_{i+k}]_{\phi} f = \int_{t_i}^{t_{i+k}} M_{i,L}(y) L^* f(y)\, dy.$$

If, in addition, $(\phi_i)_1^{k+1}$ is an extended complete Chebyshev (or, ECT) system, then Burchard [21] and Karlin [47] have shown the analog of the Schoenberg-Whitney Theorem 7.1 that, <u>for strictly increasing $\underline{\underline{t}}$ and strictly increasing $\underline{\underline{\tau}} = (\tau_i)_1^n$, $\det(M_{j,L}(\tau_i)) \geq 0$ with strict inequality iff $M_{i,L}(\tau_i) \neq 0$, all</u> i. Further, Karlin [47] showed that $(M_{j,L}(\tau_i))$ <u>is totally positive</u> in this case, as was mentioned earlier. Few facts beyong these are known for Chebyshev B-splines. While the analog of Marsden's identity (5.7) can be found in [61], the analog of the linear functional (5.4) has not been described, although that should be fairly easy. More importantly for computations, a recurrence relation like (4.9) has been searched for in vain so far.

It is actually quite unnecessary to assume that $(\phi_j)_1^k$ is a Chebyshev system in order to construct L-splines (in the sense of Greville) of local support. Continue to assume that $(\phi_j)_1^k$ is a basis for the kernel of the differential operator L^* of (9.4) with L of (9.5) its adjoint and G the Green's function given by (9.6). If $\underline{\underline{t}} = (t_i)_1^n$ is strictly increasing, then, for each i, the span of $([t_j])_{j=i}^{i+k}$ contains a nontrivial $\mu \perp \ker L^*$ since $\ker L^*$ has dimension k. But then

$$(9.10)\quad M_{\mu,L}(x) := \mu\, G(\cdot,x)$$

defines an L-spline with knots $t_i,\ldots,t_{i+k}$ and support in (t_i,t_{i+k}). Clearly, $M_{\mu,L}$ represents μ with respect to the pairing $\langle f,g\rangle := \int f L^* g$. If now $(\phi_j)_1^k$ fails to be a Chebyshev

system on $[t_i, t_{i+k}]$, then there exists a nontrivial μ in the span of $([t_j])_i^{i+r}$ and orthogonal to $\ker L^*$ for some $r < k$, i.e., the corresponding $M_{\mu,L}$ has even smaller support. More explicitly, let $(\mu_i)_1^{n+2k-2}$ be the sequence

$$[t_1]D^{k-1}, \ldots, [t_1]D, [t_1], [t_2], \ldots, [t_n], [t_n]D, \ldots, [t_n]D^{k-1}$$

of linear functionals and, for each i, let ν_i be the linear functional of the form $\nu_i = \mu_i + \Sigma_{i+1}^{i+r}\beta_j\mu_j$ which is orthogonal to $\ker L^*$, with r as small as possible. The corresponding sequence $(M_{\nu_i,L})_1^{n+k-2}$ of basic L-splines is then a basis for the space of all L-splines on $[t_1, t_n]$ with simple interior knots $t_2, \ldots, t_{n-1}$.

A construction like this was used by Jerome [45] under the additional assumption that, for each i, $t_{i+k} - t_i$ is small enough so that $(\phi_j)_1^k$ is a Chebyshev system on $[t_i, t_{i+k}]$. Earlier, Jerome and Schumaker [46] had used such considerations in connection with Lg-splines, i.e., when the linear functionals (μ_i) above are, more generally, of the form $\mu_i = \Sigma_1^k \alpha_{ij}[t_i]D^{j-1}$. Related developments of great generality can be found in Brown [20].

We close this section with yet another B-spline property discovered by Curry and Schoenberg [30].

LEMMA 9.1 [30]. Let $M_{0,k}$ be the B-spline defined by (3.1), and let σ be any k-simplex in $\mathbb{R}^k$ of unit volume with vertices $v^{(i)}$, $i = 0, \ldots, k$ and so that $v_1^{(i)} = t_i$, $i = 0, \ldots, k$. Then, for all x,

$$M_{0,k}(x) = |\sigma \cap \{v \in \mathbb{R}^k : v_1 = x\}|,$$

i.e., $M_{0,k}(x)$ gives the (k-1)-dimensional volume of the intersection of the simplex σ with the hyperplane in $\mathbb{R}^k$ which intersects the v_1-axis at $v_1 = x$ and is orthogonal to it.

In a letter [72] to P. Davis, Schoenberg recalls the Hermite-Genocchi formula

$$(9.11)\quad [z_0,\ldots,z_k]f = \int_{\tau_n}\!\!\ldots\int f^{(k)}(v_0z_0+v_1z_1+\ldots+v_kz_k)\,dv_1\ldots dv_k$$

with $v_0 = 1 - v_1 - \ldots - v_k$ and where the integration is to be carried out over the complex

$$\tau_n : v_1 \geq 0, \ \ldots, \ v_k \geq 0, \ \sum_1^k v_i \leq 1,$$

and points out that Lemma 9.1 follows from this on comparison with (4.1). Schoenberg further recalls that the Hermite-Genocchi formula remains valid if $z_0,\ldots,z_k$ are points in the complex plane not all on one line and if f is a complex-valued function regular in the convex hull $\prod$ of $z_0,\ldots,z_k$. The formula (4.1) now becomes

$$(9.12)\quad [z_0,\ldots,z_k]f = \int_{\prod} M(x,y;z_0,\ldots,z_k)\,f^{(k)}(x,y)\ dxdy/k!\ .$$

At the point $z = (x,y)$, $M(x,y;z_0,\ldots,z_k)$ is therefore the $(k-2)$-dimensional volume of the intersection of the plane $\{v \in \mathbb{R}^k : v_1 = x,\ v_2 = y\}$ with a simplex of unit volume whose i-th vertex $v^{(i)}$ satisfies $(v_1^{(i)},v_2^{(i)}) = z_i$. In particular, M is positive on $\prod$ and zero off $\prod$ and is a spline of order k-1 along any straight line, with knots only at the points where such a line intersects a segment $[z_i,z_j]$. Schoenberg's letter even contains a drawing of such a B-spline in two variables for $k = 4$.

This suggests the following definition.

DEFINITION. <u>Let</u> σ <u>be a nontrivial simplex in</u> $\mathbb{R}^{s+k}$. <u>On</u> $\mathbb{R}^s$, <u>define the B-spline of order</u> k <u>from</u> σ <u>by</u>

$$M_{k,\sigma}(x_1,\ldots,x_s) := |\sigma \cap \{v \in \mathbb{R}^{s+k} : v_i = x_i,\ i=1,\ldots,s\}|$$

all $x \in \mathbb{R}^s$.

Then $M_{k,\sigma}$ is unimodal, nonnegative, piecewise polynomial of total order k, and in C^{k-1} in general. Its support is the projection of σ onto $\mathbb{R}^s$, i.e., the convex hull of the projections $((v_j^{(i)})_{j=1}^s)_{i=0}^k$ of the vertices of σ to $\mathbb{R}^s$.

At this point, I have no idea how useful these B-splines might be, even only for the writing of papers. It is easy to visualize how such B-splines can be made to give a partition of unity: One takes some suitable convex set C in $\mathbb{R}^k$ of unit volume and then subdivides the cylinder $\mathbb{R}^s \times C$ in $\mathbb{R}^{s+k}$ into nontrivial simplices. The corresponding B-splines will then add up to one. But it is unlikely that these B-splines will become very useful unless one finds some means of evaluating them such as a recurrence relation like (4.9).

In any event, I think these B-splines are very beautiful.

Acknowledgement

It is a pleasure to acknowledge the fact that this paper reflects various valuable comments by Dr. Allan Pinkus on the first draft of the paper.

References

1. Aubin, J.-P., Behavior of the error of the approximate solution of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods, Ann. Scuola Norm. Pisa 21 (1967), 599-637.

2. ______, Approximation of Elliptic Boundary-value Problems, John Wiley & Sons, New York 1972.

3. Babuška, I., Approximation by hill functions, Comment.Math. Univ. Carolinae 11 (1970), no. 4, 787-811.

4. Birkhoff, G., Local spline approximation by moments, J. Math. Mech. 16 (1967), 987-990.

5. de Boor, C., The method of projections, etc., Ph.D. thesis, Univ. of Michigan, Ann Arbor, Mich., 1966.

6. _______, On local spline approximation by moments, J. Math. Mech. 17 (1968), 729-736.

7. _______, On uniform approximation by splines, J. Approximation Theory 1 (1968), 219-235.

8. _______, On calculating with B-splines, J. Approximation Theory 6 (1972), 50-62.

9. _______, The quasi-interpolant as a tool in elementary polynomial spline theory, in Approximation Theory (G. G. Lorentz, ed.), Academic Press, New York, 1973, 269-276.

10. _______, A remark concerning perfect splines, Bull. Amer. Math. Soc. 80 (1974), 724-727.

11. _______, How small can one make the derivatives of an interpolating function? J. Approximation Theory 13 (1975), 105-116.

12. _______, A bound on the $\mathbb{L}_\infty$-norm of $\mathbb{L}_2$-approximation by splines in terms of a global mesh ratio, MRC TSR #1597, 1975. Math. Comp. 30 (1976). To appear

13. _______, On local linear functionals which vanish at all B-splines but one, in Theory of Approximation with Applications, A. Law and A. Sahney, eds., Academic Press, New York, 1976, 120-145.

14. _______, On "best" interpolation, J. Approximation Theory 15 (1976), 28-42.

15. _______, Total positivity of the spline collocation matrix, Ind. Univ. J. Math. 25 (1976). To appear.

16. _______ and G. Fix, Spline approximation by quasi-interpolants, J. Approximation Theory 7 (1973), 19-45.

17. _______, T. Lyche, and L.L. Schumaker, On calculating with B-splines, II. Integration, in Proceedings Oberwolfach Conference Approximation Theory, May 1975 (L. Collatz, ed.)

18. _______ and A. Pinkus, Backward error analysis for totally positive linear systems, MRC TSR #1620, Jan. 1976.

19. Bramble, J. H., and S. R. Hilbert, Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation, SIAM J. Numer. Anal. 7 (1970), 112-124.

20. Brown, R. C., Generalized boundary value problems and B-splines, MRC TSR #1418, 1976, U. Wisconsin, Madison, Wis.

21. Burchard, H. G., Interpolation and approximation by generalized convex functions, Ph.D. thesis, Purdue University, Lafayette, Ind., 1968.

22. ______, Extremal positive splines with applications to interpolation and approximation by generalized convex functions, Bull. Amer. Math. Soc. 79 (1973), 959-963.

23. ______, On the degree of convergence of piecewise polynomial approximation on optimal meshes II, Trans. Amer. Math. Soc., to appear.

24. Chui, C. K., E. R. Rozema, P. W. Smith, and J. D. Ward, Simultaneous spline approximation and interpolation preserving norms, Proc. Amer. Math. Soc., to appear.

25. ______, P. W. Smith, and J. D. Ward, Favard's solution is the limit of $W^{k,p}$-splines, Trans. Amer. Math. Soc., to appear.

26. Ciesielski, Z., and J. Domsta, Construction of an orthonormal basis in $C^m(I^d)$ and $W_p^m(I^d)$, Studia Math. 41 (1972), 211-224.

27. Cox, M. G., The numerical evaluation of B-splines, J. Inst. Maths. Applics. 10 (1972), 134-149.

28. Curry, H. B., Review, Math. Tables and Other Aids to Comp. 2 (1947), 167-169 and 211-213.

29. ______ and I. J. Schoenberg, On spline distributions and their limits: the Pólya distribution functions, Abstract 380t, Bull. Amer. Math. Soc. 53 (1947), 1114.

30. ______ and I. J. Schoenberg, On Pólya frequency functions. IV. The fundamental spline functions and their limits, J. d'Analyse Math. 17 (1966), 71-107.

31. Demko, S., Local approximation properties of spline projections, J. Approximation Theory, to appear.

32. DeVore, R. A., Monotone approximation by splines, SIAM J. Math. Anal., to appear.

33. Di Guglielmo, F., Construction d'approximations des espaces de Sobolev sur des reseaux en simplexes, Calcolo 6 (1969), 279-331.

34. Domsta, J., A theorem on B-splines, Studia Math. 41 (1972), 291-314.

35. Favard, J., Sur l'interpolation, J. Mathem. pures appl. 19 (105) (1940), 281-306.

36. Fisher, S. D., and J. W. Jerome, The existence, characterization and essential uniqueness of solutions of L^∞ extremal problems, Trans. Amer. Math. Soc. 187 (1974), 391-404.

37. ______ and J.W. Jerome, Spline solutions to L^1 extremal problems in one and several variables, J. Approximation

Theory 13 (1975), 73-83.

38. Fix, G. J., and G. Strang, Fourier analysis of the finite element method in Ritz-Galerkin theory, Studies Appl. Math. 48 (1969), 265-273.

39. Gaffney, P. W., The calculation of indefinite integrals of B-splines, CSS 16, AERE Harwell, Oxfordshire, England, July, 1974.

40. Glaeser, G., Prolongement extremal de fonctions differentiables, Publ. Sect. Math. Faculte des Sciences Rennes, Rennes, France, 1967. See also

41. ______, Prolongement extremal de fonctions differentiables d'une variable, J. Approximation Theory 8 (1973), 249-261.

42. Golomb, M., $H^{m,p}$-extensions by $H^{m,p}$-splines, J. Approximation Theory 5 (1972), 238-275.

43. Greville, T. N. E., Interpolation by generalized spline functions, MRC TSR #476 (1964), U. Wisconsin, Madison, Wis.

44. ______, Introduction to spline functions, in Theory and Application of Spline Functions (T. N. E. Greville, ed.), Academic Press, New York, 1969, 87-102.

45. Jerome, J. W., Uniform approximation by certain generalized splines, J. Approximation Theory 7 (1973), 143-154.

46. ______ and L. L. Schumaker, Local bases and computation of g-splines, Methoden und Verfahren der Math. Physik 5 (1971), 171-199.

47. Karlin, S., Total Positivity, Vol. 1, Stanford University Press, Stanford, Cal., 1968.

48. ______, Some variational problems on certain Sobolev spaces and perfect splines, Bull. Amer. Math. Soc. 79 (1973), 124-128.

49. ______ and J. M. Karon, A variation-diminishing generalized spline approximation method, J. Approximation Theory 1 (1968), 255-268

50. ______ and J. M. Karon, A remark on B-splines, J. Approximation Theory 3 (1970), 455.

51. ______ and A. Pinkus, Divided differences and other nonlinear existence problems at extremal points, in Studies in Spline Functions and Approximation Theory (S. Karlin, C. A. Micchelli, A. Pinkus, and I. J. Schoenberg, eds.), Academic Press, New York, 1976, 327-352.

52. _______ and W. J. Studden, Tchebycheff Systems: with Applications in Analysis and Statistics, Interscience Publishers, John Wiley, New York, 1966.

53. _______ and Z. Ziegler, Thebysheffian spline functions, SIAM J. Numer. Anal., Series B, 3 (1966), 514-543.

54. Krein, M., The L-problem in an abstract normed space, Article IV in Some Questions in the Theory of Moments by N. I. Ahiezer and M. G. Krein, Gonti, Har'kov, 1938; Engl. transl. Transl. Math. Monographs, vol. 2, Amer. Math. Soc., Providence, R. I., 1962.

55. _______ and G. Finkelstein, Sur les fonctions de Green completement non-negatives des operateurs differentiels ordinaires, Doklady Akad. Nauk SSSR 24 (1939), 220-223.

56. Lee, S. L., B-splines for cardinal Hermite interpolation, Linear Algebra and Appl. 12 (1975), 269-280.

57. Link, Reinhardt, Asymptotische Fehlerdarstellungen bei Approximationsverfahren mit Spline-Funktionen, Doctoral dissertation, Mathematische Fakultaet, University of Freiburg, Germany, 1975.

58. Louboutin, R., Sur une bonne partition de l'unite, in Le Prolongateur de Whitney, Vol. II (G. Glaeser, ed.), Universite de Rennes, Rennes, France, 1967.

59. Lyche, Tom, and L. L. Schumaker, Local spline approximation methods, MRC TSR #1417, 1974; also J. Approximation Theory 15 (1975), 294-325.

60. Mangasarian, O. L., and L. L. Schumaker, Discrete splines via mathematical programming, SIAM J. Control 9 (1971), 174-183.

61. Marsden, M. J., An identity for spline functions with applications to variation-diminishing spline approximation, J. Approximation Theory 3 (1970), 7-49.

62. _______, On uniform spline approximation, J. Approximation Theory 6 (1972), 249-253.

63. _______ and I. J. Schoenberg, On variation diminishing spline approximation methods, Mathematica (Cluj) 8 (31), (1966), 61-82.

64. Micchelli, C. A., Best L^1-approximation by weak Chebyshev systems and the uniqueness of interpolating perfect splines, J. Approximation Theory, to appear.

65. Mühlbach, G., A recurrence formula for generalized divided differences and some applications, J. Approximation Theory 9 (1973), 165-172.

66. Passow, Eli, Piecewise monotone spline interpolation, J. Approx. Theory 12 (1974), 240-241.

67. Popoviciu, T., Sur le reste dans certaines formules lineaires d'approximation de l'analyse, Mathematica (Cluj) 1 (1959), 95-142.

68. Rice, J. R., On the degree of convergence of nonlinear spline approximation, in Approximations with Special Emphasis on Spline Functions (I. J. Schoenberg, ed.), Academic Press, New York, 1969, 349-365.

69. Riesenfeld, R., Applications of B-spline approximation to geometric problems of computer-aided design, UTEC-CSc-73-126, March 1973, Computer Science, Univ. of Utah, Salt Lake City, Utah.

70. Schoenberg, I. J., Contributions to the problem of approximation of equidistant data by analytic functions, Quart. Appl. Math. 4 (1946), 45-99, 112-141.

71. ______, On trigonometric spline interpolation, J. Math. Mech. 13 (1964), 795-826.

72. ______, letter to Philip J. Davis dated May 31, 1965.

73. ______, On spline functions, in Inequalities (O. Shisha, ed.), Academic Press, New York, 1967, 255-291.

74. ______, On spline interpolation at all integer points of the real axis, Mathematica (Cluj) 10 (33), (1968), 151-170.

75. ______, The perfect B-splines and a time-optimal control problem, Israel J. Math. 10 (1971), 261-274.

76. ______, Cardinal Spline Interpolation, CBMS Vol. 12, SIAM, Philadelphia, 1973.

77. ______ and A. Sharma, Cardinal interpolation and spline functions. V: The B-splines for Cardinal Hermite Interpolation, Linear Algebra and Appl. 7 (1973), 1-42.

78. ______ and A. Whitney, On Pólya frequency functions, III: The positivity of translation determinants with an application to the interpolation problem by spline curves, Trans. Amer. Math. Soc. 74 (1953), 246-259.

79. Schumaker, L. L., Constructive aspects of discrete polynomial spline functions, in Approximation Theory (G. G. Lorentz, ed.), Academic Press, New York, 1973, 469-476.

80. Strang, G., and G. J. Fix, An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, N. J. (1973)

Carl de Boor
Mathematics Research Center
University of Wisconsin
Madison, Wisconsin 53706

Research sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

It seems a shame to waste this almost empty page, so here is yet another reference to work concerning B-splines: In

Phillips, J. L., and R. J. Hanson, Computing integrals involving B-splines by means of specialized Gaussian quadrature rules, TR #CS-73-001, Comp. Sci., Washington State U., Pullman, WA., 1973,

one finds a discussion of the procedure for generating the three-term recurrence relation for the polynomials orthogonal with respect to a B-spline as weight function, as well as the abscissae and weights for the corresponding Gauss quadrature rule. For orders 2 and 4, and for a uniform knot sequence, specific numbers are on microfiche in

Phillips, J. L., and R. J. Hanson, Gauss quadrature rules with B-spline weight functions, Math. Comp. 28 (1974), 666.

NONLINEAR APPROXIMATION

Dietrich Braess

This survey of nonlinear approximation theory is given with emphasis on the "local-global-principle". It starts with the functional analytic approach. The theory of Rice and Meinardus/Schwedt for nonlinear Chebyshev approximation are not only considered with respect to their merits for the approximation by rational functions and by exponentials. They made the way for Wulbert's approach to global analysis continued by the author. For the treatment of the exponentials these modern tools are necessary though the theory is not yet complete for $n \geq 4$. The difference between approximation with respect to L_p-norms and the sup-norm becomes obvious.

1 Introduction

This lecture is anounced as a survey of nonlinear approximation theory. In the last years there were so many activities in so many directions that it seems to be impossible to give a fairly complete survey. Therefore, I have to focus my attention to the central problems. In order to give you an impression of the progress, the problems are discussed from the viewpoint of what will be called the "local-global principle". This principle turns out to be crucial for the total development.

Because of the incompleteness of my talk and the list of references the reader is referred to the following articles with survey character [1,4,13,27,35] and bibliographies [9, 14,34].

In nonlinear approximation theory most investigations are based on the concept of best approximation: Given a non-

void set G in a normed linear space E and an element $f \notin G$, then $g \in G$ with minimal distance from f is called a nearest point or best approximation (or short: a solution). It is called a local best approximation, if it is a best approximation to f in some neighborhood of g in G.

A set $G \subset E$ is called an existence set (uniqueness set, Chebyshev set, resp.) if for each $f \in E$ there is at least one (at most one, exactly one, resp.) best approximation. For an existence set G the metric projection $P=P_G: E \longrightarrow 2^G$ is the mapping which sends each element f to the set of its best approximations. If, moreover, G is a Chebyshev set, then P may also be understood as a mapping onto G instead of 2^G.

There are only a few non-trivial Chebyshev sets. It is remarkable that one of them was found already very early by Chebyshev. He recognized that the best approximation in the set of rational functions

(1.1) $R_{\ell,r}=\{g=p/q \in C[o,1];\ p,q \text{ polynomials, } \partial p \leq \ell, \partial q \leq r\}$

is always unique, if $C[o,1]$ is endowed with the sup-norm. The rational functions also serve as a "picture-book example" for most of the recent theories. This family behaves almost as a linear one. There is only one complication, the metric projection is not continuous at f, if Pf is degenerate, i.e. if $Pf \in R_{\ell-1,r-1}$.

On the other hand the best rational approximation is not always unique, if the function space is endowed with an L_p-norm, $1<p<\infty$. This result was first given in 1961 by Efimov and Stechkin [2o] as a byproduct of the geometric theory. In this context it is worth noting that it lasted 9 years til an example for nonuniqueness was published in the thesis of Lamprecht [24]. But as is well known the cited

result is not the only one in that theory.

Because of the extraordinary role of nonlinear Chebyshev approximation, here the analytic approaches of Rice [31] and Meinardus and Schwedt [28] are more important. Even if uniqueness does not hold, it is often possible to derive a (finite) bound for the number of solutions [7,8]. For this aim new methods must be developped [5]. From this viewpoint Wolfe's recent result on the non-existence of a bound for rational L_2-approximation [36] is of great interest.

We close the lecture with a remark on non-linear interpolation.

2 Convexity of Chebyshev sets

The first steps into the study of Chebyshev sets by geometrical tools were done by Bunt [12] in 1934 and by Motzkin [30] in 1935. It lasted 15 years til a first generalization of Motzkin's result to infinite dimensional spaces was carried out. Until the middle of the last decennium Victor Klee [23], Efimov and Stechkin, Vlasov and later Brosowski and Deutsch developped the theory. Since an excellent survey has been given recently by Vlasov [35], we restrict our attention to the main results. We will try to understand why there are still so many open problems.

DEFINITION 2.1. A normed linear space E is called strictly convex, if $\|x\|=\|y\|=1$, $x\neq y$ imply $\|\frac{1}{2}(x+y)\|<1$. It is called uniformly convex if given $\varepsilon>0$ there is a $\delta>0$ such that $\|x\|=\|y\|=1$, $\|x-y\|>\delta$ imply $\|\frac{1}{2}(x+y)\|<1-\varepsilon$. A normed linear space E is smooth, if its dual space E' is strictly convex.

Motzkin's result may be stated as follows:

THEOREM 2.2. If G is a subset of a strictly convex, smooth space of finite dimension, then the following are equivalent:

1° G *is a Chebyshev set*.

2° G *is closed and convex*.

The theorem applies to Euclidean n-space. It is still an open question whether it may be extended to Hilbert spaces without further hypothesis. On the other hand it is interesting to note under which conditions the generalization is possible.

DEFINITION 2.3. *A sequence* $\{g_\nu\}$ *in a subset* $G \subset E$ *is called a minimizing sequence, if* $\lim_{\nu \to \infty} \|f-g_\nu\| = \inf\{\|f-g\|; g \in G\}$.

G *is called approximatively compact, if each minimizing sequence in* G *contains a subsequence which converges to an element in* G. G *is called boundedly compact, if the intersection of* G *with each closed ball is compact*.

Here convergence is understood in the strong topology. Analogous definitions referring to the weak topology are made in the same manner (c.f.[12,35]).

In the following a simple observation will be crucial.

LEMMA 2.4. *The metric projection onto an approximatively compact Chebyshev set is a continuous mapping*.

The proof of a generalization of Theorem 2.2 splits into two parts. The concept of a sun is used.

DEFINITION 2.5. *A sun (strict sun, resp.) is an existence set* G *in a normed linear space* E *such that for any* $f \in E$ *and for at least one* $g \in Pf$ (for all $g \in Pf$, resp.) *we have* $g \in P[g+\lambda(f-g)]$ *for* all $\lambda > 1$.

THEOREM 2.6. *Each boundedly compact Chebyshev set in a Banach space is a strict sun*.

Sketch of proof. Let $f_1 \notin G$. Choose a ball $B = B_\delta(f_1)$ with center f_1 and radius δ which is disjoint from G. The image of

the metric projection P(B) is bounded, hence relatively compact. Now, let ϕ be the mapping, which sends V into the boundary of B:

$$\phi(g)=f_1+\delta\frac{f_1-g}{\|f_1-g\|}.$$

Consequently, $\phi\circ P$ is a continuous mapping of B into a relatively compact subset. By Schauder's theorem there is a fixed point. Uniqueness of the best approximation implies that the fixed point is $Pf_1+\lambda(f_1-Pf_1)$ with $\lambda=1+\delta(\|f_1-P(f_1)\|)^{-1}$. It follows from an open-closed argument that the assertion holds for all $\lambda>1$. $\square$

LEMMA 2.7. <u>Each sun in a smooth normed linear space is convex.</u>

<u>Sketch of proof.</u> Let G be a sun. Assume that $g_1, g_2 \in G$, but for some $\alpha\in(0,1)$ we have $f=\alpha g_1+(1-\alpha)g_2\notin G$. Let $g\in P(f)$ be a best approximation, to which the definition of a sun applies. It follows from Theorem 3.1 that for i=1,2, there is a hyperplane containing g which separates f from g_i. Smoothness implies that both hyperplanes coincide. This yields a contradiction, because f lies on the straight line connecting g_1 and g_2. $\square$

By combining Theorem 2.6 and Lemma 2.7 one gets the following theorem.

THEOREM 2.8. <u>Each boundedly compact Chebyshev set in a smooth Banach space is convex.</u>

There have been several attempts to weaken the compactness assumption and to replace it at least by the continuity of the metric projection, (continuity understood as weak as possible) [17,2o,23,35]. But it seems to be impossible to abandon it completely.

Indeed, recently Dunham [19] has discovered an example of a Chebyshev set which is not a sun.

EXAMPLE 2.9. Let $E=C[o,1]$ be endowed with the uniform norm and $G=\{F[a]\ ;\ a\geq 0\}$ where

$$F(a,t)=\begin{cases}(1+1)e^{-t/a} & \text{if } a>0,\\ 0 & \text{if } a=0.\end{cases}$$

Since the mapping $a \longrightarrow F(a)$ is continuous with respect to compact convergence on the open interval $(o,1)$, the family G is compact when endowed with that topology. By standard arguments it is an existence set. Uniqueness follows from the monotonicity: $F(a,t)-F(b,t)>0$, whenever $a>b$. But G is not a sun. The zero function is optimal for $f(t)=\frac{1}{2}$ but not for $f(t)=\frac{\lambda}{2}$, $\lambda>1$. $\square$

We conclude this section with a consideration of the continuity of the metric projection from the global point of view.

If we have only the definition of continuity in mind, then it seems to be a local property. Continuity of P onto a Chebyshev set, however, establishes global properties. For example, it implies connectedness: given $g_1, g_2 \in G$ the mapping $[o,1]: \lambda \longrightarrow P[(1-\lambda)g_1+\lambda g_2]$ furnishes a connecting arc. More generally, G must be contractible; all homology groups must be trivial. From this viewpoint it is also not only a coincidence that in the central step when proving Theorem 2.8 a fixed point theorem was involved.

I belive that the gap in the functional analytic approach results from the fact that homology is a structure which is too different in nature. Note that the family in Example 2.9 is not connected with respect to the strong topology, but it is connected with respect to the weaker topology, in which

the parametrization is continous.

We note that these difficulties are not present when manifolds are investigated. So manifolds lead to an alternative approach (s. below).

3 Suns

The concept of suns has turned out to be an effective tool in nonlinear approximation; it is not only employed in connection with the problem of the preceding section. It is our aim to characterize suns in a geometrical way (s. Lemma 3.3).

THEOREM 3.1. *For an existence set* G *in a normed linear space* E *the following are equivalent*:

1^o G *is a strict sun*.

2^o *For any* $f \in E$, *any* $g_o \in P_G(f)$ *and any* $g \in G$, *the element* g_o *is also a best approximation to* f *in the convex hull of* g *and* g_o.

3^o *Given* $f \in E$ *we have* $g_o \in P_G(f)$, *if and only if for any* $g \in G$ *there is a linear functional* $\ell \in E'$ *(which may be chosen as an extreme point in the unit ball of* E'*) satisfying*

$$\|\ell\| = 1,$$
$$\ell(f-g_o) = \|f-g_o\|,$$
$$\mathrm{Re}\,\ell(g-g_o) \leq 0.$$

The condition specified in 3^o is a generalization of Kolmogorov's criterion. The proof that it is a necessary and sufficient condition for a solution in strict suns was simplified by inserting condition 2^o. The equivalence of 1^o and 2^o follows from a simple geometrical argument. On the other hand, the equivalence of 2^o and 3^o is a consequence of Singer' characterization of best approximations in convex sets via separating hyperplanes.

We emphasize that the linear functional in Kolmogorov's criterion depends on g. But strict convexity of E' implies, that the functional is already determined by $f-g_o$. Then the difference to Singer's characterization, which is restricted to convex sets, vanishes.

In the framework of this lecture we ask for characterizations of suns by internal (or intrinsic) properties. To be more precise the specification shall not refer to the approximation problem. For other characterizations the reader is refered to the literature [1,1o,11,35]. The answer is trivial, when E is a smooth space (c.f. Lemma 2.7). A characterization of the desired type for C(Q) endowed with the sup-norm was given by Brosowski.

THEOREM 3.2. <u>(Characterization of suns). A non-void set</u> $G\subset C(Q)$ <u>is a sun if given a pair</u> $g,g_o\in G$ <u>and</u> $\lambda>0$, <u>then for every closed set</u> A <u>satisfying</u>

$$\min_{t\in A} |g(t)-g_o(t)|>0$$

<u>there is a</u> $g_\lambda\in G$ <u>such that</u> $\|g_\lambda-g_o\|<\lambda$ <u>and</u>

$$\min_{t\in A} \text{Re}\left\{\overline{(g_\lambda(t)-g_o(t))}(g(t)-g_o(t))\right\}>0.$$

The above characterization shows the connection between local and global features: g may be an element which is arbitrarily far from g_o, but it has an influence on each neighborhood of g_o.

It is probably a question of personal taste as to whether the generalizations of Theorem 3.2 [1o,11] are still considered as characterizations by "intrinsic" properties.

In the finite dimensional case there is an intuitive geometrical interpretation for the characterization of suns [4].

LEMMA 3.3. Let G be a closed set in C(Q), Q being a (finite) set with m points. Then the following are equivalent:

1^{o}. G is a sun.

2^{o}. Each pair of elements g_0, $g_1 \in G$ may be connected by a continuous curve $\{g_\lambda, 0 \leq \lambda \leq 1\}$ in G with the following property: if the i-th coordinates of g_0 and g_1 are not equal, i.e., if $(g_1 - g_0)_i \neq 0$, then the i-th coordinate $(g_\lambda)_i$ is a strictly monotonic function of λ.

Figure 1 illustrates that the property of being a sun is comparable to convexity. Connecting curves exist in both contexts, but the condition on the curve in the sun situation is less restricting.

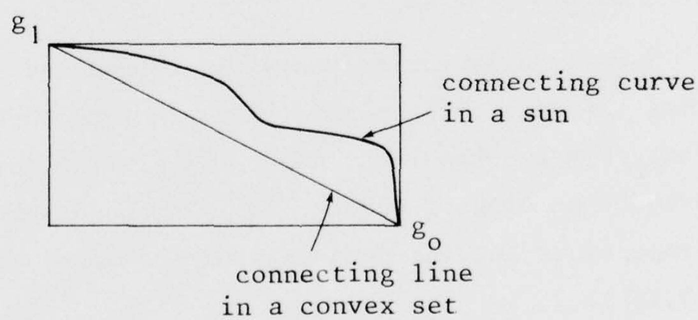


Fig. 1. Comparison of connecting curves in suns and convex sets.

4 Varisolvency

In the preceding sections we have discussed the approach in the framework of functional analysis. In 1961 J.R. Rice introduced a completely different concept for nonlinear Chebyshev approximation [31]. (It may be thought of as the first analytical approach.) Here and in the sequel I denotes a compact non-degenerate real interval.

DEFINITION 4.1. Let $G \subset C(I)$.

(i) G is said to be locally solvent of degree m at g_o, if given $\varepsilon>0$ and m distinct points $x_i \in I$, $i=1,2,\dots,m$, there is a $\delta>0$, such that $|g_o(x_i)-y_i|<\delta$, $i=1,2,\dots,m$, implies that there is a $g \in G$, satisfying $\|g-g_o\| < \varepsilon$ and

$$g(x_i)-y_i=0, \quad i=1,2,\dots,m.$$

(ii) G is said to have Property Z of degree m at g_o, if $g-g_o$ has at most m-1 zeros whenever $g \in G$, $g \neq g_o$.

(iii) G satisfies the density property, if given $g_o \in G$ and $\varepsilon>0$, there are $g_1, g_2 \in G$ such that $\|g_i-g_o\|<\varepsilon$, $i=1,2$, and

$$g_1(t)<g_o(t)<g_2(t), \quad t \in I.$$

A family G with the density property is said to be varisolvent, if with each $g \in G$ an integer m(g) can be associated such that G is locally solvent at g and satisfies Property Z both with degree m(g).

Originally varisolvency was introduced without postulating the density property (iii). In 1968 Dunham discovered that this hypothesis or something similar must be added to produce an elegant theory [18]. On the other hand the density property is not independent from the other postulates (c.f. [3,25]).

THEOREM 4.2. For each $f \in C(I)$ there is at most one best approximation in a varisolvent family G. Moreover, $g \in G$ is a best approximation, iff f-g alternates at least m(g) times.

Here, alternating is understood as in the classical theory for linear Haar subspaces.

We mention this theory not only because of its merits for applications (see below) but because of its clear structure. Observe that for the definition of varisolvency (i) is a local property whilst (ii) is a global one. We note

that the local degree cannot be larger then the global one. Rice's theory makes clear in which sense the connection between local and global behaviour is needed in the analytic approach.

When Rice's theory is applied to families of actual interest, the degrees from Definition 4.1 must be calculated. Here, the theory of Meinardus and Schwedt [28] yields an effective tool and has also prepared the way for the more modern treatment.

At this point is seems appropriate to discuss the role played by the parametrizations. In his original paper Rice formulated his results for families which are given in a parametrized form

$$G = \{F(a,.);\ a \in A\}.$$

Here A is a (locally compact) parameter set and $F: A \times I \longrightarrow \mathbb{R}$ a continuous mapping. This has led to confusion in the literature. In particular one has to distinguish between the topology of C(I) and the topology of the parameter set. However, Rice's results may be formulated and derived without any reference to parameters.

On the other hand the use of local parameters is often advantageous. It may even be considered as a typical phenomenon of nonlinear analysis that no global parametrization is taken; the choice of parametrization depends on the element in the family, which (and the neighborhood of which) is actually being considered. A byproduct of the theory of Meinardus and Schwedt [28] is the following.

LEMMA 4.3. <u>Let</u> $G \subset C(I)$. <u>Assume that</u> F <u>is mapping from an open set</u> $A \subset \mathbb{R}^m$ <u>into</u> G. <u>Moreover, let</u> F <u>be continuously differentiable in the sence of Fréchet. If the "tangent</u>

space" $d_aF(\mathbb{R}^m)$ is an m-dimensional Haar subspace of $C(I)$ then G is locally solvent of degree m at $g=F(a)$.

The lemma is easily proved by the use of the classical implicit-function-theorem.

Obviously a family G is varisolvent, if for each $g\in G$ there is a local parametrization as given in the lemma whose "dimension" m coincides with the degree $m(g)$ of Property Z at g. The density postulate causes no trouble here, because the tangent space is a Haar subspace and contains a positive function.

EXAMPLE 4.4. Let $R_{\ell,r}$, $\ell,r>0$ denote the set of rational functions given in (1.1). The defect of a function p/q is

$$\operatorname{def}\left(\frac{p}{q}\right)=\begin{cases} r & \text{if } p=0,\\ \min(\ell-\partial p, r-\partial q) & \text{if } p\neq 0.\end{cases}$$

$R_{\ell,r}$ is varisolvent and the degree at p/q is $m(g)=\ell+r+1-\operatorname{def}(g) \leq \ell+r+1$.

EXAMPLE 4.5. Let E_n^o, $n\geq 1$ denote the set of proper (sums of) exponentials:

$$E_n^o = \{g=\sum_{\nu=1}^{n} \alpha_\nu e^{\beta_\nu t}; \alpha_\nu, \beta_\nu \in \mathbb{R}\}.$$

An exponential g has the order $k=k(g)$, if k is the least number such that $g\in E_k^o$. It is shown in [2] that E_n^o is varisolvent and that the degree at g is $n+k(g)\leq 2n$.

The metric projection (if it exists) is continuous if the image point is normal in G.

DEFINITION 4.6. An element $g\in G$ is a normal point in a varisolvent family G if one of the following equivalent properties holds:

1^o The degree is constant in a neighborhood of g.

2^o There is a neighborhood of g in G, the closure of which

is compact.

3^o A neighborhood of g in G is a manifold (i.e. is homeomorphic to an open set in a Euclidean space).

There is no convention in the literature as to which property is to be used in other contexts than the theory of varisolvency. To complete the confusion we emphasize that Condition 1^o is not equivalent to the statement that the degree is maximal. This is shown by an example.

EXAMPLE 4.7. Let $G=G_1 \cup G_2$ where

$$G_1 = \{g(t)=\alpha, \quad \alpha \leq 0\},$$
$$G_2 = \{g=\alpha e^{\beta t}, \quad \alpha>0, \ \beta \in \mathbb{R}\}.$$

G is a connected set. It is varisolvent of degree m at g, if $g \in G_m$, $m=1,2$. Apart from $g=0$ all elements are normal in G.

5 Critical point theory. Manifolds.

The theory of varisolvent families requires a very strong global assumption (Property Z) on zeros. This is due to the fact that the families need not be manifolds. Indeed, the rational functions in Example 4.4 and the exponentials in Example 4.5 do not form manifolds. It is the aim of this section to weaken the global assumptions as much as possible. The first step into this direction was done in 1971 by Wulbert [37,38].

To illustrate the basic concept we consider a simple example (c.f. Figure 2). Let G be a smooth curve in 2-space satisfying the following local property: the gradient shall be positive (or negative, resp.) at each point. It follows from this fact already that there is a parametrization such that each coordinate is a strictly monotonic function. The curve is a uniqueness set for the Chebyshev approximation.

Fig. 2: Curve with nonzero gradient in $\mathbb{R}^2$.

The generalization to manifolds with a dimension greater than one is by no means trivial. (Compare with the fact that Brouwer's fixed point theorem is trivial only for the dimension one, because in that case it reduces to the intermediate value theorem.)

The appropriate framework is critical point theory [5].

DEFINITION 5.1. Let G be a non-void subset in a normed linear space R, and $g \in G$. Then the tangent cone $C_g G$ at g to G consists of all elements $h \in R$ with the following property: there is a continuous mapping of $[0,1]: \lambda \longrightarrow g_\lambda \in G$ such that $g_0 = g$ and $\|g_\lambda - g - \lambda h\| = o(\lambda)$ as $\lambda \longrightarrow 0$.

DEFINITION 5.2. Let $G \subset R$ be non-void and $f \in R$. Then $g \in G$ is called a critical point, if 0 is a best approximation to $f-g$ in $C_g G$.

The importance of these definitions becomes obvious from the following:

LEMMA 5.3 (Lemma of first variation). If g is a local best approximation to f in G, then g is a critical point.

Proof. Assume that g is not a critical point. Then there is a tangent vector $h \in C_g G$ such that

$$C := \|f-g\| - \|f-g-h\| > 0.$$

We compute the distance of f from the elements of the curve

given in Definition 5.1:

$$\begin{aligned}\|f-g_\lambda\| &\le \|f-g-\lambda h\| + \|g_\lambda - g - \lambda h\| \\ &\le \lambda\|f-g-h\| + (1-\lambda)\|f-g\| + o(\lambda) \\ &\le \|f-g\| - c\lambda + o(\lambda) < \|f-g\|,\end{aligned}$$

provided that λ is sufficiently small. Hence, g is not locally optimal. $\square$

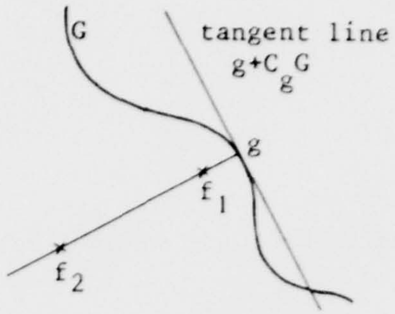
Note that the main step in the proof above is nothing else than an application of Newton's method.

The standard application of the lemma is the following. One tries to find a linear subspace of $C_g G$, e.g. by looking for an appropriate parametrization. Then the criteria from linear theory are applied. In particular, in the Chebyshev case the Kolmogoroff-criterion may be used [28].

The tangent cones are linear sets, if G is a manifold without boundary.

DEFINITION 5.4. A subset $G \subset R$ is an m-dimensional C^1-submanifold of R, if for each $g_0 \in G$ there is a neighborhood $U \subset G$ and a one-one mapping F from an open set $W \subset \mathbb{R}^m$ onto U. Moreover, F is assumed to be continuously Fréchet differentiable on W, and $d_a F$ to be injective for each $a \in W$.

It is crucial for critical point theory that the converse of Lemma 5.3 does not hold in general. But it may be proved for the situation given in Theorem 5.6. At first we regard Fig. 3 which illustrates two different situations in 2-space. In the left figure g is a critical point to f_1 and f_2, since the connecting line is orthogonal to the tangent line. However, g is a local best approximation only with respect to f_1 and not to f_2. The right figure refers to the sup-norm. Then g is a critical point and a local best approximation to f. It is a local solution to each element on the



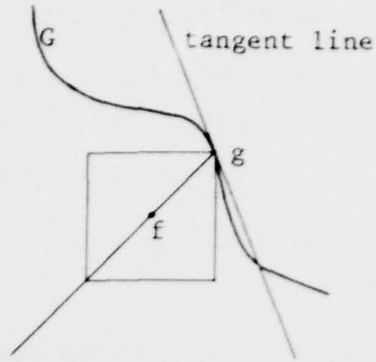


Fig. 3. Critical points, in the situation of Euclidean norm and sup-norm.

ray through f with center g.

To put this observation into a rigorous theory the following concept was introduced by Wulbert [37].

DEFINITION 5.5. A C^1-submanifold $G\subset C(I)$ is Haar embedded, if all tangent spaces (tangent cones) are Haar subspaces of $C(I)$.

THEOREM 5.6. Let $G\subset C(I)$ be a Haar embedded manifold and let $f\in C(I)$. Then g is a local best approximation to f in G, iff g is a critical point.

Proof. Because of Lemma 5.1 it is only necessary to prove the "if"-part.

Assume that g_o is a critical point to f in G. Let $F:W \longrightarrow G$ be a parametrization such that $F(o)=g_o$. The parametrization induces a mapping from $F(W)$ into $C_{g_o} G$,

$$g \longrightarrow h = d_o F(F^{-1} g).$$

It follows from the differentiability that

$$\| g-g_o-h \| = o(\| h \|).$$

By the definition of critical points 0 is a best approximation to $f-g_o$ in the Haar subspace $C_{g_o} G$. As is known from

linear theory, then 0 is also a strong best approximation. This means that we have for some $c>0$:

$$\|f-g_o-h\| \geq \|f-g_o-o\| + c\|h\|,$$

whenever h is a tangent vector. Combining both inequalities we obtain

$$\begin{aligned}\|f-g\| &\geq \|f-g_o-h\|-\|g-g_o-h\| \\ &\geq \|f-g_o\|+c\|h\|-o(\|h\|) \geq \|f-g\|,\end{aligned}$$

if $\|g-g_o\|$ is sufficiently small. □

Now we turn to the development of the global result. As usual, for $\alpha \in \mathbb{R}$ the level set

$$G^\alpha = \{g \in G; \|f-g\| \leq \alpha\}$$

is introduced.

THEOREM 5.7 <u>(Deformation Theorem). Let</u> G <u>be a</u> C^1<u>-manifold. If the set</u> $\{g \in G;\ \alpha \leq \|f-g\| \leq \beta\}$ <u>is compact and contains no critical point, then</u> G^α <u>is a strong deformation retract of</u> G^β.

<u>Outline of proof.</u> Consider a non-critical point g_o. Let F be a parametrization for some neighborhood U of g_o in G. Since g_o is not critical, there is a parameter vector b such that

$$g_o = F(a_o), \qquad \|f-g_o-d_{a_o}F\, b\| < \|f-g\|.$$

With the same arguments as in the proof of Lemma 5.3 we get

$$\|f-F(a+\lambda b)\| < \|f-F(a)\|$$

for all parameters a in some neighborhood of a_o and sufficiently small λ. We may choose a cut-off-function κ such that $\kappa(a_o)>0$ and

$$\|f-F(a+\lambda\kappa(a)\cdot b)\| \leq \|f-F(a)\|, \qquad 0 \leq \lambda \leq 1,$$

whenever $F(a) \in U$. Consequently, by the transformation

$$g \xrightarrow{F^{-1}} a \longrightarrow a+\lambda\kappa(a)\cdot b \xrightarrow{F} g_\lambda$$

a flow

$$\phi:[0,1]\times G \longrightarrow G$$

has been constructed such that

$$\phi(0,g)=g,$$
$$\|f-\phi(\lambda,g)\|<\|f-g\|,\; 0<\lambda\leq 1.$$

In particular, the inequality is strict when $g=g_o$.

The proof of the theorem proceeds by covering the set $\{g\in G,\ \alpha\leq\|f-g\|\leq\beta\}$ by a finite number of open sets such that for each point at least one of the corresponding flows reduces the distance to f. By glueing the flows together and using compactness once more, we obtain for each $g\in G^\beta$ a path of descent, which ends in G^α. $\square$

As a consequence we have [5,38]:

THEOREM 5.8 (Uniqueness Theorem). Each connected, boundedly compact, Haar embedded submanifold of $C(I)$ is a Chebyshev set.

REMARK. Haar embeddednes may be considered as an assumption on zeros with local structure. Assumptions referring to global properties are only connectedness and boundedly compactness.

Proof. Assume that there are two local best approximations g_1 and g_2 in G. Since G is connected, they belong to the same component of G^β, if β is sufficiently large. The set of critical points is closed, hence compact if restricted to G^β. Consequently, we may choose two local solutions g_3, g_4 (which possibly are g_1 and g_2) in the component such that there is no further critical point between the levels $\alpha=\max\{\|f-g_i\|,\ i=3,4\}$ and β. However, the local solutions g_3 and g_4 must be

contained in distinct components of $G^{\alpha+\varepsilon}$, ε sufficiently small. This contradicts the deformation theorem. $\blacksquare$

6 Approximation by sums of exponentials

In non-linear Chebyshev approximation the rational functions and the exponentials are the families which are the most investigated. Whilst the rational functions show a good behaviour, the treatment of exponentials is very involved.

The set E_n^o of proper exponentials given in Example 4.5 is not an existence set. In order to assure existence one has to consider the closure of E_n^o.

$$(6.1)\quad E_n=\{g=\sum_{\nu=1}^{\ell} p_\nu(t)e^{\beta_\nu t}\ ;\ k=\sum_{\nu=1}^{\ell}(1+\partial p_\nu)\leq n,\ \beta_\nu\in\mathbb{R}\}.$$

Here, the p_ν's are polynomials with degree ∂p_ν. If an (extended) exponential is represented as in (6.1), then $\ell=\ell(g)$ corresponds to the number of distinct characteristic numbers, while $k=k(g)$ is evaluated with counting multiplicities. The first rigorous proof for E_n being an existence set was given by Werner in 1968. The reader is referred to E. Schmidt's paper [33] for an easier proof.

As was observed in 1968 [2] the extended family E_n is not varisolvent für $n\geq 2$. It is locally solvent of degree $n+\ell(g)$ at g, while Property Z holds with degree $n+k(g)$. Note the gap between ℓ and k, if $g\in E_n\setminus E_n^o$. Nevertheless, local best approximations may be characterized by an alternating of the error curve [8]. The number of alternatings lies between $n+\ell$ and $n+k$.

But we have not always uniqueness. By using Meinardus' invariance principle examples for functions with two best approximations were constructed [2]. For example, to $f(x)=(1+x^2)^{-1}$ on $I=[-1,+1]$ there are two best approximations in E_2. On the other hand it was also possible to prove with classical

tools that there are at most two solutions in E_2. For several years, however, it was not known whether the number of solutions in E_n, $n\geq 3$, is bounded.

This problem was attacked by using the tools from global analysis described in the last section. Three years ago the author anounced the result that there are at most n! local solutions in E_n. Unfortunately, there was still a gap in the proof. Anyway, in the meantime the case n=3 could by settled [8]. The number of local solutions in E_3 is less or equal than 3, and this bound is the best possible. (More generally, the bound for E_n cannot be less than n.) We hope that within the next years the little gap in the proof for $n\geq 4$ can be closed by the development of appropriate perturbation techniques.

The results on Haar embedded manifolds may be applied. But we have to overcome two difficulties.

At first the theory of Haar embedded manifolds must be extended to include manifolds with boundaries. Then the tangent cones are no longer linear spaces but proper cones. Fortunately, this causes only complications, which are technical in nature. They are treatable though they make the analysis a littly clumsy.

More annoyance is caused by the fact that E_n is not a manifold. This becomes already obvious from the standard representation for E_1.

$$\begin{aligned} \mathbb{R}^2 &\longrightarrow E_1 \\ (\alpha,\beta) &\longrightarrow \alpha e^{\beta t} \end{aligned}$$

This parametrization is only one-one, if we contract the line $\{(0,\beta),\beta\in\mathbb{R}\}$ in $\mathbb{R}^2$ to a single point. In the resulting parameter set zero is a singular point. Note that the singular

point corresponds to an exponential which has not the maximal order. If we want to obtain a manifold, then we have to consider $E_n \setminus E_{n-1}$ instead of E_n.

Refering to E.Schmidt's compactness results for exponentials, we obtain from Theorem 5.8:

LEMMA 6.1. Each component of $(E_n)^\alpha$, $\alpha \in \mathbb{R}$, which is disjoint from E_{n-1}, contains exactly one local best approximation.

Now, consider a local best approximation in E_n and the components of the corresponding level sets. There is a smallest α such that the level set with level α intersects E_{n-1}. A central result of the investigation is that the minimal intersection contains an element which is a local solution with respect to E_{n-1}. Consequently, we can characterize the solutions in E_n by those in E_{n-1}.

For enumerating the solutions in E_n via an inductive process the converse problem arises: how many solutions in E_n can bifurcate from any solution $\hat{g}$ in E_{n-1}? To be more precise, consider the component of the level set $(E_n)^\alpha$, $\alpha = \| f - \hat{g} \|$, which contains $\hat{g}$. Into how many parts will it be splitted, if the level decreases by a small number ε to $(E_n)^{\alpha-\varepsilon}$?

Let us return for a moment to the discussion of the singular point in E_1. From the viewpoint of differential geometry it is clear how to cancel the singularity. One has to blow up the point to a 1-dimensional line. The question arises whether this blowing up can be done consistent with the analysis.

Indeed, $\hat{g} \in E_{n-1}$ may be written as an exponential in E_n:

$$\hat{g} + 0 \cdot e^{\beta t},$$

where elements with different β are not identified. Now descend from $\hat{g} + \alpha_o e^{\beta t}$, $\alpha_o = 0$, β fixed, by Newton's method to a

lower level. It turns out that two choices β and β' are equivalent if no characteristic number of $\hat{g}$ is contained in the (closed) interval with endpoints β and β'. Since $\hat{g}$ has at most n-1 characteristic numbers there are at most n inequivalent choices of β. (There is still some trouble for the neighborhoods of the spectrum of g). By an induction the bound n! for the number of possibilities is conjectured.

7 Nonlinear mean-squares approximation

One of the first results on rational approximation with respect of L_p-norms, $1<p<\infty$, was the result of Cheney and Goldstein [15] that each (local) best approximation in $R_{\ell,r}$ has zero defect. Therefore, one of the complications in rational Chebyshev approximation does not happen here. The non-uniqueness result mentioned in the introduction, remembers us, however, that we cannot expect a nice theory. We will return to the non-uniqueness problem later.

At first we will have a glance at the general theory. During the last years the lemma of first variation (Lemma 5.3) was (re-) discovered by several authors. A deeper analysis is possible if we restrict our attention to L_2, or more generally to a Hilbert space H with inner product $[\ ,\ ]$. It turns out that most investigations are based on the following analysis of second order terms.

Let A be an open set in n-space, and let $F:A \longrightarrow G$ be a parametrization. Assume that the first and the second derivative d_aF and d_a^2F in the sense of Fréchet exist. Here, d_aF is a linear transformation: $\mathbb{R}^n \longrightarrow H$ and d_a^2F is a bilinear symmetric form: $\mathbb{R}^n \times \mathbb{R}^n \longrightarrow H$. If the kernel ker d_aF consists only of the zero element, then $C_{F(a)}G = d_aF(\mathbb{R}^n)$.

The analysis is based on the square of the distance function

$$\rho(a)=\| f-F(a)\|^2=[f-F[a],f-F(a)].$$

The derivatives of ρ are easily derived.

$$d_a\rho=-2[f-F(a),d_aF]$$
$$\frac{1}{2}d_a^2\rho=[d_aF,d_aF]-[f-F(a),d_a^2F]$$

An element $g=F(a)$ is a critical point, if $d_a\rho=0$. Concerning second order terms the following definitions are useful [29].

DEFINITION 7.1. Let $F(a)$ be a critical point. The number of negative eigenvalues of $d_a^2\rho$ is called its index, and the dimension of $\ker d_a^2\rho$ is called its nullity. A critical point is degenerate, if its nullity is greater then zero.

Note that the classical criteria say the following: If a critical point is a local solution, then its index is zero. On the other hand each critical point with vanishing index and nullity is a local best approximation.

While the meaning of the index is obvious, it is not so quickly understood for the nullity. Non-degenerate critical points are isolated. Moreover, the location of a non-degenerate critical point is a continuous function of f (Compare the problem of the continuity for the metric projection). The most important result incorporating nullity is obtained in connection with a solar property.

The expression for $d_a^2\rho$ splits into two terms in a natural way. The term $[d_aF,d_aF]$ is called the first fundamental form; the other one $[F(a)-f,d_a^2F]$ is the second fundamental form.

If $F[a]$ is a critical point to f, then it is also critical, when the element

$$f_\lambda=F(a)+\lambda(f-F(a)),\ \lambda>0$$

is to be approximated. The corresponding second derivative is

$$[d_aF, d_aF] - \lambda[f-F(a), d_a^2F].$$

Hence, only the second fundamental form depends on λ. Therefore, the determination of the degenerate situations corresponds to an eigenvalue problem. If $F(a)$ is a degenerate critical point to f_λ, then f_λ is called a focal point.

LEMMA 7.2. (Morse's index theorem) The index of a non-degenerate critical point g is equal to the number of focal points which lie on the segment from f to g; each focal point being counted with its multiplicity.

The index vanishes for sufficiently small λ. This corresponds to the fact that the curvature (which is defined by the n eigenvalues mentioned above) is finite.

The situation for nonlinear uniform approximation is different. Chui and Smith [16] observed recently that a critical point in a smooth manifold need not be a local solution, even if the element to be approximated is arbitrarily close. Another example is given in Fig. 4. In the light of Rice's definition of (a scalar) curvature [32], one might say that the curvature is infinite. But this definition has a serious disadvantage: the curvature is here not an intrinsic property, it depends on the norm.

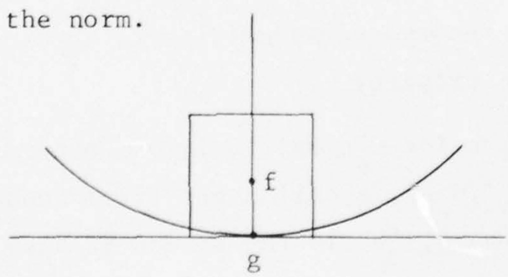


Fig. 4. Example for critical points which are not local best approximations, when the tangent line does not satisfy the Haar condition.

The splitting of the second derivatives into the two fundamental forms is also used for the proof the following theorem which was first given by Wolfe [36] for the case $\ell < r$.

THEOREM 7.3. Assume that $p_i/q_i \in R_{\ell,r} \setminus R_{\ell-1,r-1}$, $\partial q_i = r$, $i=1,2,\ldots,N$, are such that q_i and q_j have no common factors unless $i=j$. Then there are a $f \in L_2$ and N elements $g_i \in R_{\ell,r}$ which differ from p_i/q_i by a polynomial and which are local best approximations to f.

REMARK. One has $g_i = p_i/q_i$ if $\ell < r$.

Sketch of proof. The main idea of the proof is to solve the equations

$$\left.\begin{aligned} [f-g_i,\ d_{a_i}F] &= 0 \\ [f-g_i,\ d_{a_i}^2 F] &= 0 \end{aligned}\right\} \quad i=1,2,\ldots,N.$$

The first set of equations assures that all g_i's are critical points, while the second set causes that the second fundamental forms vanish. Since the first fundamental form is definite, all g_i's are local best approximations.

The derivatives d_aF and d_a^2F do not change, if a polynomial is added. Consequently, the equations above are linear equations for f and $(g_i - p_i/q_i)$, $i=1,2,\ldots,N$. The existence of a solution is established after expressing the derivatives in terms of appropriate rational functions. □

From the theorem it follows immediately that there is no uniform bound for the number of local best approximations in $R_{\ell,r}$. Therefore, the situation is even worse than in the Chebyshev approximation by exponentials.

8 Approximation by interpolation

Up to now we have always considered the concept of best approximation. One can also try to find a good approximation by interpolating at some points. However, there does not always exist a solution when one wants to interpolate by rational functions, by exponentials or by spline functions with free knots. Curiously, the conditions which garantee existence, are very similar in the three examples.

At first let us focus our attention to rational interpolation [6]. The following interpolation problems was already considered by Cauchy in 1823. Given $\ell+r+1$ points $(t_i, y_i) \in \mathbb{R}^2$, $i=0,1,\ldots,\ell+r$, the function $p/q \in R_{\ell,r}$ is an interpolant, if

$$\frac{p(t_i)}{q(t_i)} = y_i, \qquad i=0,1,\ldots,\ell+r.$$

If p/q interpolates, then we have

$$p(t_i)-y_i q(t_i)=0, \qquad i=0,1,\ldots,\ell+r.$$

This system of $(\ell+r+1)$ linear homogeneous equations for $(\ell+r+2)$ polynom-coefficients has always a solution. But the solution may be only a formal solution. It may have poles between the given t_i's.

The following result generalizes a result of Löwner [26].

THEOREM 8.1. Let $\ell \geq r-1$ and $I=[-1,+1]$. Assume that $f \in C(I)$ has an integral representation

$$f(t)=p(t)+\int_{-1}^{+1} \frac{t^{\ell-r+1}}{1-tx}\, d\mu(t),$$

where p is a polynomial of degree $\ell-r$ and $d\mu$ is a non-negative measure with infinite carrier. Then given $t_i \in I$, $i=0,1,\ldots,\ell+r$, there is a $g \in R_{\ell,r}$ which interpolates f at the

given points. Moreover, the interpolation error is pointwise smaller when compared with the interpolation by polynomials of degree $(\ell+r)$.

For exponentials or spline functions with free knots similar results are obtained if the (iterated) Cauchy kernel $t^m(1-xt)^{-1}$ in the representation is replaced by the exponential kernel $\exp(tx)$ or the truncated power kernel $(t-x)_+^n$, resp. The connection of the interpolation problems with moment problems is very obvious in the treatment of the exponential case in [22], while the approach in [21] shows that homotopy-techniques are appropriate tools, when the abscissa are not equally spaced.

References

1 Amir, D. and F. Deutsch, Suns, moons and quasipolyhedra, J. Approximation Theore, 6 (1972), 176-2o1.

2 Braess, D., Approximation mit Exponentialsummen. Computing, 2 (1967), 3o9-321.

3 -, On varisolveney and alternation. J. Approximation Theory, 9 (1973), 2o-43.

4 -, Geometrical characterization for nonlinear uniform approximation, J. Approximation Theory, 11 (1974), 26o-274.

5 -, Kritische Punkte bei der nichtlinearen Tschebysheff-Approximation. Math. Z., 132 (1973), 327-341.

6 -, Rationale Interpolation, Normalität und Monosplines. Numer. Math., 22 (1974), 219-232.

7 -, On the number of best approximations in certain non-linear families of functions. aequationes math., 12 (1975), 184-199.

8 -, Chebyshev approximation by γ-polynomials, III. (in preparation).

9 Braess, D., J. Brink-Spalink und R. Schaback, Eine Bibliographie zur nichtlinearen Approximationstheorie. Report, Universität Bonn 1974.

10 Brosowski, B. and F. Deutsch, On some geometrical properties of suns. J. Approximation Theory, 1o (1974), 245-267.

11 Brosowski, B. and R. Wegmann, Charakterisierung bester Approximationen in normierten Vektorräumen. J. Approximation Theory, 3 (197o), 369-397.

12 Bunt, L., Bijdrage tot de theorie der convexe puntverzamelingen. Thesis univ. Groningen 1934.

13 Cheney, E.W., A survey of recent progress in approximation theory. aequations math. (to appear).

14 -, A bibliography for approximation theory, 197o-1974. Report CNA-94, University of Texas, Austin, 1974.

15 Cheney, E.W. and A.A. Goldstein, Mean-square approximation by generalized rational functions. Math.Z., 95 (1967), 232-341.

16 Chui, C.K. and P.W.Smith, Unique best nonlinear approximation in Hilbert spaces. Proc. Amer. Math. Soc. 49 (1975), 66-78.

17 Deutsch, F. and J.M. Lambert, The geometry of Banach spaces with regard to metric projections and approximative compactness. (submitted)

18 Dunham, C.B., Necessity of alternation. Canad. Math. Bull, 11 (1968), 743-744.

19 -, Chebyshev sets in C[o,1] which are not suns. Canad. Math. Bull. (to appear).

2o Efimov, N.V. and S.B. Stechkin, Approximative compactness and Chebyshev sets. Dokl. Akad. NAUK SSSR, 140 (1961) 522-524 = Soviet Math. Dokl. 2 (1961), 1226-1228.

21 FitzGerald, C. and L.L. Schumaker, A differential equation approach to interpolation at extremal values. J. d'Analyse Math., 22 (1969), 117-134.

22 Karlin, S., Total Positivity. Stanford: Stanford University Press 1968.

23 Klee, V.L., Convexity of Chebyshev sets. Math. Ann., 142 (1961), 292-3o4.

24 Lamprecht, G., Zur Mehrdeutigkeit bei der Approximation in der L_p-Norm mit Hilfe rationaler Funktionen. Computing, 5 (1970), 349-355.

25 Ling, W.H. and J.E. Tornga, The constant error curve problem for varisolvent families. J. Approximation Theory, 11 (1974), 54-72.

26 Löwner, K., Über monotone Matrixfunktionen. Math. Z., 38 (1933), 177-216.

27 Lorentz, G.G., Monotone approximation. In "Inequalities III", edit. by O.Shisha. pp. 2o1-215. Academic Press: New York 1972.

28 Meinardus, G. und D. Schwedt, Nichtlineare Approximation. Arch. Rational Mech. Anal., 17 (1964), 297-326.

29 Milnor, J., Morse Theory, Princeton: University Press 1963.

3o Motzkin, T.S., Sur quelques propriétés caractéristiques des ensembles convexes. Atti Accad. Naz. Lincei Rend. Ser. 4, 21 (1935), 562-567.

31 Rice, J.R., Tchebycheff approximation by functions unisolvent of variable degree. Trans. Amer. Math. Soc., 99 (1961), 298-3o2.

32 -, Nonlinear approximation II. Curvature in Minkowski geometry and local uniqueness. Trans. Amer. Math. Soc., 128 (1967), 437-459.

33 Schmidt, E., Zur Kompaktheit bei Exponentialsummen. J. Approximation Theory, 3 (197o), 445-454.

34 Taylor, G.D., Uniform approximation with side conditions. In "Approximation Theory" edit. by G.G. Lorents, pp. 495-5o4. Academic Press: New York-London 1973.

35 Vlasov, L.P., Approximative properties of sets in normed linear spaces. Uspehi Mat. NAUK 28, 6 (174), (1973), 3-66 (Russian) = Russian math. surveys, 28, 6, 1-66.

36 Wolfe, J.M., On the unicity of nonlinear approximation in smooth spaces. J. Approximation Theory, 12 (1974), 165-181.

37 Wulbert, D.E., Uniqueness an differential characterization of approximation from manifolds of functions. Amer. J. Math., 93 (1971), 35o-366.

38 -, Nonlinear approximation with tangential characterization, Amer. J. Math., 93 (1971), 718-73o.

D. Braess
Mathematisches Institut
der Ruhr-Universität
Bochum
463 Bochum, Universitätsstr. 15o
Germany (West)

RECENT RESULTS ON PADÉ APPROXIMANTS AND RELATED PROBLEMS

Charles K. Chui

The purpose of this paper is to study the recent development of certain aspects of Padé approximants and related problems. For instance, convergence results of the approximants and their connection with best local approximation will be discussed.

1. Introduction.

The subject of Padé approximants is fairly old. It dates back to as early as Cauchy (1789-1857) and Jacobi (1804-1851) but was first treated in detail by Frobenius in 1881 [63]. Under the influence of Hermite (1822-1901) at the Ecole Normale Supérieure, Padé, in his 1892 dissertation [127], classified these rational fraction approximants (now known as Padé approximants), arranged them in a table (which is now called the Padé table) and studied the structure of it in detail. In those early days, Padé approximants were considered as another form of continued fraction, which was already a well established subject. Indeed, if

$$f(z) = b_0 + \cfrac{a_1 z}{1 + \cfrac{a_2 z}{1 + \ddots}} = c_0 + c_1 z + c_2 z^2 + \ldots ,$$

and $P_0(z) = b_0$, $P_1(z) = b_0 + a_1 z$, $P_2(z) = b_0 + a_1 z/(1+a_2 z)$, . . . are the truncations of the above continued fraction, then the even ones $P_0, P_2, P_4, \ldots$ occupy the main diagonal and the odd ones $P_1, P_3, P_5, \ldots$ march down the $[n+1/n]$

diagonal of the Padé table of the formal power series $c_0 + c_1 z + c_2 z^2 + \ldots$. References to the subject of Padé approximants and continued fractions can be found in the books Baker [7], Balk [12], Luke [105, 106], Perron [131] and Wall [157], and Proceedings of conferences [10], [84] and [92], as well as the articles by Gragg [81] and Wynn [175].

During the early 1960's, physicists found that Padé approximants are very useful as a systematic method of extracting more information from (formal) power series expansions, and the method has been especially successful in critical phenomena. In scattering theory and quantum field theory where series of Stieltjes occur frequently, the approximants are known to converge to the expected solutions rapidly. However, when the series do not have the "positive character", much still has to be understood. The main contributors in this direction include the physicists or chemists Baker, Bessis, Chisholm, Fisher, Gammel, Nuttall, Wheeler, and others. The interested reader should consult the articles listed in the references. It should be noted, however, that hundreds of papers related to Padé approximants have been written during the past twelve years, and Graves-Morris of the Mathematical Institute, University of Canterbury, Kent, England, is preparing a bibliography.

In the next section, we will discuss the definition of Padé approximants and several generalizations. The third section of this paper will be devoted to the important question of convergence. In the fourth or final section, the related problems of best local approximation and best quasi-rational approximation will be treated.

2. The Padé approximants.

The definition of Padé approximants is a natural extension of that of the Taylor series. Let

$$(2.1) \quad f(z) = \sum_{k=0}^{\infty} c_k z^k$$

be a formal power series with complex coefficients c_k, and let m and n be non-negative integers. The $[m/n]$ Padé approximant of f (at the origin) is the unique rational function $[m/n] \equiv [m/n](f) \equiv p_m/q_n$ where $q_n \not\equiv 0$ and

$$p_m(z) = \sum_{k=0}^{m} a_k z^k \quad \text{and} \quad q_n(z) = \sum_{k=0}^{n} b_k z^k$$

such that

$$f(z)q_n(z) - p_m(z) = d\, z^{m+n+1} + \text{higher order terms.}$$

The existence of p_m and q_n follows immediately from considering the system of equations

$$(2.2) \quad \sum_{k=0}^{j} c_{j-k} b_k = \begin{cases} a_j & \text{for } j = 0,\dots, m \\ 0 & \text{for } j = m+1,\dots,m+n \end{cases}$$

(where we set $b_k \equiv 0$ if $k > n$), and the uniqueness of the ratio p_m/q_n is an easy exercise. Hence, the following table, which is called the Padé table, is obtained:

$$\begin{array}{llll} [0/0] & [1/0] & [2/0] & \dots \\ [0/1] & [1/1] & [2/1] & \dots \\ [0/2] & [1/2] & [2/2] & \dots \\ \dots & \dots & \dots & \dots \end{array}$$

It is clear that the first row is actually the sequence of partial sums of the formal series f itself. We include some other useful equivalent definitions in the following theorem.

THEOREM 2.1. Let f be the formal power series (2.1) and let f_{m+n} be the $(m+n)^{th}$ partial sum of f. Then each of the following statements is equivalent to the definition of Padé approximants $[m/n] = p_m/q_n$ of f:

(a) $f(z)q_n(z) - p_m(z) = s\, z^{\mu}$ + higher order terms, where $\mu \leq \infty$ is as large as possible and $q_n \not\equiv 0$.

(b) $f(z) - p_m(z)/q_n(z) = t\, z^{\nu}$ + higher order terms, where $\nu \leq \infty$ is as large as possible and $q_n \not\equiv 0$.

(c) $(f_{m+n}\, q_n - p_m)^{(j)}\,(0) = 0$ for $j = 0,\ldots, m+n$, and $q_n \not\equiv 0$.

Statement (b) can be found in Padé [127]. We remark that ν may happen to be smaller than $m+n+1$, although it is clear that $\mu \geq m+n+1$. However, if the series (2.1) is normal (cf. Frobenius [63]), that is, if

$$(2.3)\qquad \begin{vmatrix} c_u & c_{u-1} & \cdots & c_{u-v+1} \\ c_{u+1} & c_u & \cdots & c_{u-v+2} \\ c_{u+v-1} & c_{u+v-2} & \cdots & c_u \end{vmatrix} \neq 0 \quad ,$$

where $c_w \equiv 0$ if $w < 0$, for all u and $v \geq 0$, then it can be shown that $\nu = m+n+1$ also. The normality condition is a very important one in the theory of Padé approximants; for example, when f is normal, the fractions p_m/q_n are in irreducible form for all m and n, so that the Padé table of f has no blocks of order larger than one (cf. Gragg [81]). If f is a function of continuity class C^{m+n} at the origin, then we can use (c) as the definition of the $[m/n]$ Padé approximant of f as in Chui, Shisha and Smith [34] by replacing f_{m+n} by f in (c), namely,

$$(2.4)\qquad (f\, q_n - p_m)^{(j)}\,(0) = 0\ ,\quad j = 0,\ldots, m+n.$$

We call (2.4) the Padé equations in [34]. One advantage of this definition is that we do not have to write down the Taylor expansion of f at 0.

It is intuitively clear from the definition of Padé approximants that they are good approximants at the origin. It is also clear that they are quite computable if, for example, the coefficients c_k of the formal power series are known (or can be estimated). Many algorithms have been written; we only mention Claesson [37] and Gragg [81] and refer the reader to the references at the end of this paper.

Next, we mention several important generalizations of the Padé approximants:

For each k, we let t_k be a polynomial of degree k and leading coefficient equal to 1, so that $t_0 \equiv 1$, and consider the formal polynomial series

$$(2.5) \quad f(z) = \sum_{k=0}^{\infty} c_k\, t_k(z).$$

Suppose that there exist polynomials p_m and q_n, $q_n \not\equiv 0$, with degrees m and n respectively, such that

$$(2.6) \quad f(z)q_n(z) - p_m(z) = c\, t_{m+n+1}(z) + d\, t_{m+n+2}(z) + \ldots ,$$

we say that p_m/q_n is a generalized $[m/n]$ Padé approximant of f. The difficulty in proving existence here is that we do not have the simple relationship $t_m\, t_n = t_{m+n}$ (as in the case of a formal power series with $t_n(z) = z^n$), so that the equations (2.2) are not satisfied.

For instance, let $\{\beta_k\}$, $k = 1,2,\ldots$, be a preassigned sequence of (interpolation) points in the complex plane and let

$$(2.7) \quad t_0(z) \equiv 1, \quad t_k(z) = \prod_{j=1}^{k} (z-\beta_j) \quad \text{for } k \geq 1 .$$

Then (2.5) is called a formal Newton series. It can be proved in this case (cf. Gallucci, Jones [64], Karlsson [95, 96] and Warner [164, 165]) that polynomials p_m and q_n which satisfy (2.6) exist and that the ratio p_m/q_n is unique. Here, p_m/q_n is called the $[m/n]$ Newton-Padé approximant of f in (2.5). In the special case if the interpolation points β_k are distinct, and the Newton series (2.5) converges to f uniformly on a set containing the points β_k, then p_m/q_n is the "best" interpolating rational function of f at $\{\beta_k\}$ of degree (m,n) in the sense that if $r_{m,n}$ is any rational function of degree (m,n) such that $r_{m,n}$ interpolates f in at least the same subset of $\{\beta_1,\ldots,\beta_{m+n+1}\}$ as p_m/q_n then we have $r_{m,n} \equiv p_m/q_n$. The more general case when we have

$$t_0(z) \equiv 1, \quad t_k(z) = \prod_{j=1}^{k} (z-\beta_j^{(k)}) \quad \text{for} \quad k \geq 1 ,$$

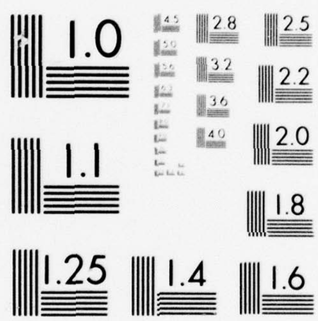
where $\{\beta_j^{(k)}\}$, $j = 1,\ldots,k$, $k = 1,2,\ldots,$ are points in the complex plane, has also been studied by Karlsson [95, 96], and Warner [164, 165]. In this case, assuming f to be of continuity class C^{m+n} in an "open" set containing the points $\{\beta_j^{(k)}\}$, $j = 1,\ldots,k$, $1 \leq k \leq m+n+1$, we have to define p_m and q_n in a slightly different way, namely, instead of (2.6), they should satisfy

$$(2.8) \quad (f\, q_n - p_m)\,(\beta_j^{(k)}) = 0 \,, \quad q_n \not\equiv 0, \; j = 1,\ldots,k \,, \quad \text{and} \quad k = 1,\ldots,m+n+1,$$

where for each k, derivatives of order μ, $\mu = 0,\ldots,\nu-1$, have to be taken if ν of the points $\beta_j^{(k)}$ coincide. Of course, if all of the $\beta_j^{(k)}$ are equal to 0, then (2.8) becomes the set of Padé equations as in (2.4). We call

AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1
APPROXIMATION THEORY II
NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.
UNCLASSIFIED AFOSR TR-76-1242 N/L

2 OF 7
AD
A033032



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

p_m/q_n the best interpolating rational function of degree (m,n) to f at $\{\beta_j^{(k)}\}$ (or an $(m+n+1)$-point Padé approximant).

Now suppose that the polynomials t_k are normalized Chebyshev polynomials, that is

$$t_0 \equiv 1, \quad t_k(x) = 2^{-k+1} \cos k \cos^{-1} x \ , \ x \in [-1,1], \ k \geq 1.$$

Then (2.5) is called a formal Chebyshev series. In this case, existence and uniqueness of the generalized Padé approximants have been studied by Clenshaw and Lord [40] and Maehly [109]. Also, Gragg and Johnson [83] rewrote the Chebyshev series as a formal Laurent series and hence studied the so-called Laurent-Padé approximants. Other formal series of orthogonal polynomials have been considered by Holdaman [89].

Another natural and at least computationally useful generalization is to consider formal power series of two variables. The research in this direction is quite recent. It has been initiated by Chisholm [30] in 1973, and then extended in 1974 and 1975 by the Canterbury group (cf. [86], [133] and the references there). The approximants are now called Canterbury approximants. The idea of this extension is that if one of the two variables is kept constant, then one should get the one-variable Padé approximants.

The concept of Padé approximants can also be generalized in the following way. Let $f_1, \ldots, f_N$ $(N \geq 2)$ be formal power series (of the same variable) as in (2.1). Determine polynomials $p_1, \ldots, p_N$, of degrees $n_1, \ldots, n_N$ respectively, such that not all of these polynomials are identically zero and

(2.9) $f_1(z)p_1(z) + \ldots + f_N(z)p_N(z) = a\, z^{n_1+\ldots+n_N+N-1} +$ higher order terms.

This problem was first posed by Padé [129] in 1894, and later has been studied in detail by Jager [90] and de Bruin [26] in their 1964 and 1974 Amsterdam theses respectively. Of course if $N = 2$ and $f_2 \equiv -1$, (2.9) becomes the original Padé problem. Let $N = 3$, $f_3 \equiv 1$, $f_2 \equiv f$ and $f_1 \equiv f^2$; then we obtain the quadratic Padé approximants as considered by Shafer [144]. Joyce and Guttmann (cf. Gammel [67]) recently considered the special case where $f_1 \equiv f$, $f_2 \equiv f', \ldots, f_N \equiv f^{(N-1)}$. This special case is important in physical problems concerning critical phenomena, and it is especially important when $N = 3$, since for the case where p_1, p_2, p_3 are restricted to polynomials of degrees 6, 8 and 9 respectively it is related to the Onsager solution for the free energy of an Ising model on a square lattice.

Other important generalizations include operator Padé approximants. An interested reader should consult the articles listed in the references.

3. Convergence of Padé approximants.

The problem of convergence of a certain sequence of Padé approximants is in general a very difficult but certainly interesting and important one. The choice of a suitable sequence from the Padé table depends very much on the analytic character of the function that the (formal) power series represents. As one might expect, the convergence proofs usually rely, either directly or indirectly, on some knowledge of the location of the poles of the approximants; and finding information on the location of the poles

is a very important unsolved problem in the theory of Padé approximants. We divide our discussion of the convergence problem into the following four categories.

I. Convergence along rows. For a fixed positive integer j, the j^{th} row on the Padé table is the sequence $\{[n/j-1]\}$, $n = 0, 1, 2, \ldots$. Hence, if f is a holomorphic function in the disc

$$\Delta_R = \{z : |z| < R\} , \quad R > 0 ,$$

then the first row, being the sequence of partial sums of the Taylor expansion of f at 0, certainly converges uniformly on every compact subset of Δ_R to f. Suppose again that f is holomorphic in Δ_R. The problem of convergence along other rows is much harder. In fact, Perron [131; p. 270] gave an example of an entire function f such that the poles of the $[n/1]$ Padé approximants of f form a dense set in the complex plane. Hence, one has to go to subsequences. In [15], Beardon indeed proved that if f is holomorphic in Δ_R, then there is a subsequence of the sequence of $[n/1]$ Padé approximants that converges uniformly on every compact subset of Δ_R to f. Very recently, Baker and Graves-Morris [11] proved that the third row $[n/2]$ has the same property. But nothing else is known for the other rows. However, if f has a finite number of poles in Δ_R, then there is the following famous classical result of de Montessus [119].

THEOREM 3.1 *Let f be holomorphic in Δ_R except for N poles $z_1,\ldots,z_N$ with a total multiplicity $M \geq N$ such that $z_1,\ldots,z_N \neq 0$. Let $\Delta_R' = \Delta_R \backslash \{z_1,\ldots,z_N\}$. Then the sequence $\{[n/M]\}$, $n = 1,2,\ldots,$ of Padé approximants of f converges uniformly on every compact subset of Δ_R' to f as $n \to \infty$.*

Wilson [168] has extended the above result to the convergence along the $(M+\mu)^{th}$ rows $(\mu > 1)$ for certain "smooth" functions f which even include the cases when f may have non-polar singular points on the boundary of Δ_R (cf. [7, p. 143]). The result of de Montessus was also proved for $(n+M+1)$ - point Padé approximants by Saff [138], which lead to the recent work of Karlsson [94, 95, 96] and Warner [164, 165] on the general interpolation problem as mentioned in section 2. As a consequence of Theorem 3.1, it is not difficult to show (cf. Wallin [158]) that if f is holomorphic in Δ_R except for infinitely many poles at $K = \{z_1, z_2, \ldots\}$, such that $0 \notin K$, then there exist subsequences $\{m_j\}$ and $\{n_j\}$ of the sequence of positive integers such that the $[m_j/n_j]$ Padé approximants of f converge, as $j \to \infty$, uniformly on each compact subset of $\Delta_R \backslash K$ to f.

II. Convergence along the main diagonal.

The problem of convergence of the sequence $\{[n/n]\}$ of Padé approximants, which occupy the main diagonal of the Padé table, is especially interesting. However, except in the case of a series of Stieltjes (and related ones) it has been difficult to obtain satisfactory results. We first state the following conjecture given by Baker, Gammel and Wills (cf. [6; p. 23 - p. 24]).

CONJECTURE 3.1 *Let* f *be holomorphic in* $\bar{\Delta}_1 = \{z : |z| \leq 1\}$ *except for a finite number of poles in* $0 < |z| < 1$ *and except for* $z = 1$ *where* f *is continuous relative to* $\bar{\Delta}_1$. *Then there exists a subsequence of the sequence of* $[n/n]$ *Padé approximants of* f *that converges to* f *uniformly on every compact subset of* $\bar{\Delta}_1 \backslash K$, *where* K *is the set of poles of* f.

No proof and counterexamples to this conjecture are known. Perhaps the first breakthrough in convergence along the main diagonal is the following result of Nuttall [121].

THEOREM 3.2. Let ℓ be any fixed integer. If f is a meromorphic function in the whole complex plane and does not have a pole at the origin, then the sequence $\{[n + \ell /n]\}$, $n = 1,2,\ldots,$ of Padé approximants of f converges in measure to f as $n \to \infty$.

Pommerenke [132] extended the above result of Nuttall to convergence in capacity for a slightly larger class of functions, namely those holomorphic in the whole complex plane except for a set of capacity zero which does not contain the origin. Pommerenke considered sequences $[m/n]$ with $c^{-1}n \le m \le c\, n$, $c > 1$. Recently, Gončar ([79] and cf. [95]) has generalized the results of Nuttall and Pommerenke still further. Another important result is the following theorem due to Wallin [159].

THEOREM 3.3. Let $\alpha > 0$ and $\{n_k\}$ be a subsequence of the sequence of positive integers. Suppose that $f(z) = \Sigma\, a_j z^j$ is holomorphic in Δ_R and satisfies

$$(3.1) \qquad \sum_{k=1}^{\infty} \left(\max\{|a_j| : n_k < j \le 2n_k\}\right)^{\alpha/n_k} < \infty .$$

Then the sequence $\{[n_k/n_k]\}$ of Padé approximants of f converges uniformly on each compact subset of $D\backslash E$ to f as $k \to \infty$, where E is a set of α - dimensional Hausdorff measure zero.

In the above, E is a set such that given any $\varepsilon > 0$, it can be covered by the union of countably many discs with radii r_k such that $\Sigma\, r_k^{\alpha} < \varepsilon$. For $\alpha = 2$, E is a set of (two-dimensional) Lebesgue measure zero. We remark that

the above theorem was actually stated for a sequence $[m_k/n_k]$ in [159] under a little different condition, and that the theorem is sharp in the sense that for each $\alpha > 0$, and any set E with α - dimensional Hausdorff measure zero not containing the origin, there exist an entire function f and a sequence $\{n_k\}$ such that (3.1) is satisfied and such that the sequence of Padé approximants $[n_k/n_k]$ of f is divergent, and even unbounded, everywhere on E. We take $\alpha > 2$, then since the whole complex plane is of α - dimensional Hausdorff measure zero, we obtain the following corollary.

COROLLARY 3.1 <u>There is an entire function</u> f <u>such that the sequence of</u> $[n/n]$ <u>Padé approximants of</u> f <u>is divergent everywhere in the whole complex plane except at the origin</u>.

As mentioned previously, the difficulty in a convergence proof is that it is in general not easy to have any control of the poles of the approximants. In fact, Wallin [159] has proved that for any sequence of points $z_k \neq 0$, $k = 1,2,\ldots$, in the complex plane, and of integers n_k with $n_{k+1} > 2n_k$, there exists an entire function for which the $[n_k/n_k]$ Padé approximant has a pole at z_k, $k = 1,2,\ldots$. Perhaps the following theorem due to Jones and Thron [93] which relates the convergence and boundedness of a sequence of Padé approximants is of some interest.

THEOREM 3.4 <u>Let</u> $\{m_k\}$ <u>and</u> $\{n_k\}$ <u>be sequences of non-negative integers such that</u> $\Sigma\, \varepsilon^{m_k} < \infty$ <u>for some</u> ε, $0 < \varepsilon < 1$. <u>Let</u> f <u>be a formal power series with Padé approximants</u> $[m_k/n_k]$ <u>and</u> D <u>a domain containing the origin</u>. <u>Then the sequence</u> $\{[m_k/n_k]\}$ <u>converges uniformly on each compact subset of</u> D <u>if and only if for all</u>

sufficiently large k, the approximants $[m_k/n_k]$ are uniformly bounded on every compact subset of D.

For a series of Stieltjes which will be discussed next, we do have some control of the poles (and actually know where they lie) for certain approximants.

III. Series of Stieltjes. The only simple general class of formal power series whose "diagonal" Padé approximants are known to converge nicely is the class of series of Stieltjes. A formal power series

$$(3.2) \quad f(z) = \sum_{j=0}^{\infty} c_j z^j$$

is called a series of Stieltjes if there is a real-valued, bounded, nondecreasing function μ assuming infinitely many values on $[0, \infty)$ such that

$$(3.3) \quad c_j = \int_0^{\infty} t^j d\,\mu(t) \quad , \quad j = 0, 1, 2, \dots \quad .$$

It was Stieltjes [148] who first proved that the moment problem on $[0, \infty)$ associated with the sequence $\{c_j\}$ is determinate if and only if the corresponding continued fraction expansion of f converges everywhere in the complex plane except on $[0, \infty)$. Since the truncations of the continued fraction expansion of f are the $[n/n]$ or $[n+1/n]$ Padé approximants of f (cf. section 1), it follows that if, say,

$$(3.4) \quad \sum_{j=1}^{\infty} c_j^{-1/2j} = \infty$$

(cf. Carleman [30]; for instance, $c_j = O((2j)!\, R^{2j})$ for some $R > 1$), then each of the sequences $\{[n/n]\}$ and $\{[n+1/n]\}$ converges to the Stieltjes integral

$$(3.5)\quad \int_0^\infty \frac{d\mu(t)}{1 - zt}$$

uniformly on every compact set in the complex plane which is disjoint from $[0,\infty)$, the positive half of the real axis. This result can also be obtained by using determinant theory [7, 11], Schwinger variational principle [1, 3], theory of orthogonal polynomials [97], or by just observing that the $[n+\ell/n]$, $\ell \geq -1$, Padé approximants are certain Gaussian quadratures of the integral (3.5) (cf. [2]). We summarise some useful explicit results, which also give error estimates, in the following theorem (cf. [3, 97]).

THEOREM 3.5 *Let* $f(z) = \Sigma c_j z^j$ *be a series of Stieltjes with measure* $d\mu$ *as described above and let* $\ell \geq -1$. *Suppose that*

$$L_n(z) \equiv L_{n,\ell}(z) = \sum_{j=0}^{n} b_{n,j} z^j$$

are the orthogonal polynomials with respect to $\{[0,\infty)\,;\ t^{\ell+1} d\mu(t)\}$ *with leading coefficients* $b_{n,n} = b_n$, *zeros at* $x_{n,k}$ *and Christoffel numbers* (*or Gaussian weights*) $\alpha_{n,k}$, $k = 1,\ldots, n$. *Then the* $[n+\ell/n]$ *Padé approximant of* f *is given by*

$$[n+\ell/n](z) \equiv \frac{P_{n+\ell}(z)}{Q_n(z)} = \sum_{j=0}^{\ell} c_j z^j + z^{\ell+1} \sum_{k=1}^{n} \frac{\alpha_{n,k}}{1 - z\,x_{n,k}}$$

which is the Gaussian quadrature approximation to

$$\sum_{j=0}^{\ell} c_j z^j + z^{\ell+1} \int_0^\infty \frac{t^{\ell+1} d\mu(t)}{1 - zt} \equiv \int_0^\infty \frac{d\mu(t)}{1 - zt}\ ,$$

where

$$P_{n+\ell}(z) = \sum_{k=0}^{n-1} z^k \sum_{j=n-k}^{n} b_{n,2n-j-k}\, c_{n-j}\ , \quad \text{and}$$

$$Q_n(z) = \sum_{k=0}^{n} b_{n,n-k}\, z^k \quad .$$

The error is given by the following formulae

$$E_n(z) \equiv E_{n,\ell}(z) \equiv \int_0^\infty \frac{d\mu(t)}{1-zt} - [n+\ell/n]\ (z)$$

$$= \frac{z^{\ell+1}}{L_n(z^{-1})} \int_0^\infty \frac{L_n(t)\ t^{\ell+1}}{1-zt}\ d\mu(t)$$

$$= \frac{z^{2n+\ell+1}}{z^n\ L_n(z^{-1})} \int_0^\infty \frac{L_n(t)\ t^{n+\ell+1}}{1-zt}\ d\mu(t)$$

$$= \frac{z^{\ell+1}}{L_n^{\,2}(z^{-1})} \int_0^\infty \frac{L_n^2\,(t)\ t^{\ell+1}}{1-zt}\ d\mu(t) \quad .$$

The poles of $[n+\ell/n]$ are simple and located on $[0,\infty)$; and on $(-\infty,\ 0)$, we have

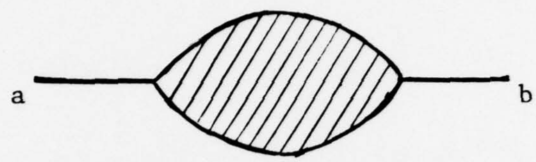
$$(-1)^{\ell+1}\ \{[n+\ell+1/n+1] - [n+\ell/n]\} > 0\ ,$$

$$(-1)^{\ell+1}\ \{[n+\ell+1/n-1] - [n+\ell/n]\} < 0\ , \quad \text{and}$$

$$(-1)^{\ell+1}\ E_n > 0.$$

Hence, the convergence result follows immediately if $\ell \geq -1$. Analogous results for the case $\ell < -1$ seem to be unknown. From the above theorem, we only have convergence along the diagonals $[n+\ell/n]$, $\ell \geq -1$, parallel to the main

diagonal of the Padé table. From a recent communication by Pommerenke, we learnt that Stahl [147] has recently proved that if $m = c\,n$, where c is some positive constant, then the sequence of $[m/n]$ Padé approximants of a series of Stieltjes of f whose corresponding measure $d\mu$ has support on a bounded interval $[a,b]$, converges, as $n \to \infty$, uniformly on each compact set disjoint from the following set:



If the series (3.2) comes from an integral

$$\int_C \frac{d\mu(t)}{1-zt}$$

where $d\mu > 0$ and C is an arc in the complex plane, analogous convergence results can still be obtained (cf. Nuttall [125]). Recently, Nuttall ([126] and private communication) has obtained results in a much more general case where the weights are allowed to be complex-valued.

IV. Convergence for certain functions. As mentioned above, for series that are not those of Stieltjes, or related series considered by Nuttall and others, it has been difficult to obtain general convergence results. In Balk [12; p. 249 - p. 257] and Perron [131 ; p. 246 - p. 248], convergence results have been obtained for the functions

$$e^z, \quad \sum q^{n^2} z^n, \quad a + \sum \frac{q^n}{1+q^n} z^n \quad ,$$

$$a + \sum \frac{z^n}{\Gamma(b+n+1)} \quad \text{and} \quad \prod (1-q^n z) .$$

More recently, convergence results for the Padé approximants of other functions including certain ratios of trigonometric polynomials [51], functions of the form $\Sigma\, \alpha_n/(1-\beta_n z)$ [60, 61, 69], quasi-meromorphic functions (where a finite number of essential isolated singular points are allowed) [50], functions of the form $a\, e^{bz} \prod (1 - \alpha_j z)/(1 - \beta_j z)$ [5], and ratios of certain special functions (cf. [105, 106, 107]) have been obtained. In particular, Padé approximants of the important function e^z have been studied in detail by Saff and Varga [139, 140] and Saff, Varga and Ni [142]. Of special interest is that the optimal sequence of approximants from the Padé table of e^{-z} which minimizes the error on $[0,\infty)$ is the sequence $\{[n/3n]\}$.

4. Best local approximation.

In 1934, Walsh [160] noted that the Taylor polynomial $\Sigma_{k=0}^{n} a_k z^k$, which is the $[n/0]$ Padé approximant, of a function holomorphic at the origin could be obtained by taking the limit, as $\varepsilon \to 0^+$, of the net of n^{th} degree polynomials which best approximate f in the closed discs $\bar{\Delta}_\varepsilon$ under the uniform norm. Later [161, 162], he proved the following result for other Padé approximants.

THEOREM 4.1 *Let* f *be in* $C^{m+n+1}[0,\delta]$ *for some* $\delta > 0$ *and write*

$$(4.1)\quad f(x) = a_0 + a_1 x + \dots + a_{m+n} x^{m+n} + O(x^{m+n+1})$$

as $x \to 0^+$. Let $r_\varepsilon \equiv r_\varepsilon(f)$ *denote a rational function of degree* (m,n) *which best approximates* f *in the uniform norm on the closed interval* $[0, \varepsilon]$, $0 < \varepsilon \le \delta$. *Suppose that the determinant*

$$(4.2)\qquad \begin{vmatrix} a_m & a_{m-1} & \cdots & a_{m-n+1} \\ a_{m+1} & a_m & \cdots & a_{m-n+2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m+n-1} & a_{m+n-2} & \cdots & a_m \end{vmatrix} \neq 0 \quad \underline{\text{and}} \quad a_0 \neq 0$$

where $a_j \equiv 0$ if $j < 0$. Then as $\varepsilon \to 0^+$, the net $\{r_\varepsilon\}$ converges to the $[m/n]$ Padé approximant of f, uniformly on any closed interval $[0, \varepsilon_0]$, $0 < \varepsilon_0 \le \delta$, on which $[m/n] \equiv [m/n](f)$ is analytic.

The normality condition (4.2) of f is needed in Walsh's proof to guarantee $||r_\varepsilon - f||_{[0,\varepsilon]} = O(\varepsilon^{m+n+1})$ which follows from the fact that $||[m/n] - f||_{[0,\varepsilon]} = O(\varepsilon^{m+n+1})$ (cf. section 2). In [34], Chui, Shisha and Smith used an approximation theoretic proof to obtain the following result, with the normality condition dropped.

THEOREM 4.2. Let f be a real-valued function in $C^{m+n+1}[0,\delta]$ for some $\delta > 0$, and for each ε, $0 < \varepsilon \le \delta$, let r_ε be the rational function of degree (m,n) which best approximates f on $[0,\varepsilon]$ in the uniform norm. Then the net $\{r_\varepsilon\}$ converges uniformly to the $[m/n]$ Padé approximant of f on some closed interval $[0,\varepsilon_0]$, $0 < \varepsilon_0 \le \delta$, as $\varepsilon \to 0^+$.

The above mentioned results lead to the concept of best local approximation which we describe as follows: Let M be a class of functions in $C[0, \delta]$, $\delta > 0$. Suppose that for each $\varepsilon > 0$ with $0 < \varepsilon \le \delta$, a function $f \in C[0, \delta]$ has a best uniform approximant $p_\varepsilon(f)$ on $[0, \varepsilon]$ from M; that is

$$||p_\varepsilon(f) - f||_{[0,\varepsilon]} = \inf\{||p - f||_{[0,\varepsilon]} : p \in M\}.$$

If as $\varepsilon \to 0^+$, the net $\{p_\varepsilon(f)\}$ converges uniformly on some closed interval $[0,\varepsilon_0]$, $0 < \varepsilon_0 \le \delta$, to some function $p_0(f) \in M$, we say that $p_0(f)$ is a best local approximant of f from M. Hence, if M is the class of all rational functions of degree (m,n), the above theorem says that for every (real-valued) function $f \in C^{m+n+1}[0,\delta]$, $\delta > 0$, the $[m/n]$ Padé approximant of f is the unique best local approximant of f from M. For $u_1,\ldots,u_n \in C[0,\delta]$, we denote by $S_n \equiv S_n(u_1,\ldots,u_n)$ the algebraic span of $\{u_1,\ldots,u_n\}$. In [35], Chui, Shisha and Smith obtained the following result.

THEOREM 4.3. Let $\{u_1,\ldots,u_n\} \subset C^n[0,\delta]$ for some positive δ, such that S_n is a Haar subspace. For each $f \in C^n[0,\varepsilon]$, $0 < \varepsilon \le \delta$, let $p_\varepsilon(f)$ denote the best uniform approximant of f on $[0,\varepsilon]$ from S_n. Then the net $\{p_\varepsilon(f)\}$ converges, as $\varepsilon \to 0^+$, uniformly on some $[0,\varepsilon_0]$, $0 < \varepsilon_0 < \delta$, to some function $p_0(f) \in S_n$ for each $f \in C^n[0,\varepsilon]$ if and only if the Wronskian matrix at 0

$$(4.3)\quad A_n \equiv \begin{bmatrix} u_1(0) & u_2(0) & \cdots & u_n(0) \\ \cdot & \cdot & \cdots & \cdot \\ u_1^{(n-1)}(0) & u_2^{(n-1)}(0) & \cdots & u_n^{(n-1)}(0) \end{bmatrix}$$

is nonsingular. Furthermore, if $\det A_n \neq 0$, we have

$$(p_0(f) - f)^{(j)}(0) = 0 \quad , \quad j = 0,\ldots,n-1 .$$

In [36], Chui, Smith and Ward considered the problem of best local approximation using the $L_2[0,\varepsilon]$ norm. There, by using a different method, better results can be obtained under some weaker continuity conditions.

Now suppose that the matrix A_n in (4.3) is singular; it seems to be an interesting problem to study what functions $f \in C^n[0,\delta]$ have the property that the nets $\{p_\varepsilon(f)\}$ of

best approximants would converge as $\varepsilon \to 0^+$. To study this problem, the notion of Taylor rank was used in [35]. Let $u_1,\dots,u_n \in C[0,\delta]$ be sufficiently smooth. The smallest integer N such that the $N \times n$ matrix

$$(4.4)\quad U_N = \begin{bmatrix} u_1(0) & u_2(0) & \cdots & u_n(0) \\ \cdot & \cdot & \cdots & \cdot \\ u_1^{(N-1)}(0) & u_2^{(N-1)}(0) & \cdots & u_n^{(N-1)}(0) \end{bmatrix}$$

has rank n is called the Taylor rank of the system $\{u_1,\dots,u_n\}$. (We require at least $u_1,\dots,u_n \in C^N[0,\delta]$ and remark that $N \geq n$). Hence, if $\det A_n \neq 0$, then $N = n$ and $U_N = A_n$. Let Ω_N be the image of R^n, the real Euclidean n-space, under the transformation U_N. The following result was obtained in [35].

THEOREM 4.4. <u>Let</u> N <u>be the Taylor rank of</u> $\{u_1,\dots,u_n\} \subset C^N[0,\delta]$ <u>and let</u> $f \in C^N[0,\delta]$ <u>where</u> $\delta > 0$. <u>Suppose that the</u> N-<u>vector</u>

$$\hat{f} \equiv (f(0)\ ,\ \dots\ ,\ f^{(N-1)}(0)\ )$$

<u>lies in</u> Ω_N. <u>Then the net</u> $\{p_\varepsilon(f)\}$ <u>of best uniform approximants of</u> f <u>from</u> S_n <u>converges to some</u> $p_0(f) \in S_n$ <u>as</u> $\varepsilon \to 0^+$. <u>Furthermore</u>,

$$(p_0(f) - f)^{(j)}(0) = 0 \quad \underline{\text{for}} \quad j = 0,\dots,N-1\ .$$

We remark that the Haar condition is not required in the above theorem. Next, since Padé approximants are defined by rational functions p_m/q_n with $(f\,q_n - p_m)(x) = 0(x^{m+n+1})$, it seems natural to consider approximation pairs (p, q) such that $fq - p$ is "smallest". More precisely, let P and Q be finite dimensional subspaces of $C[0,\delta]$, $\delta > 0$. Assume that $\dim P = m$, $\dim Q = n$ and $\dim Q_0 = n - 1$ where

$$Q_0 = \{q \in Q : q(0) = 0\} .$$

Let $Q_1 = Q_0 + 1$, and for each $f \in C[0,\varepsilon]$, $0 < \varepsilon \leq \delta$, consider the minimization problem

$$(4.5) \quad \inf \{||fq - p||_{[0,\varepsilon]} : p \in P , q \in Q_1\} .$$

The minimizing pairs $(p_\varepsilon , q_\varepsilon)$ can easily be shown to exist. If the subspace

$$R_f \equiv P + fQ_0$$

is Haar on $[0,\varepsilon_0]$, then for each ε, $0 < \varepsilon \leq \varepsilon_0$, the minimizing pair $(p_\varepsilon, q_\varepsilon)$ for f is also unique. Let $\{p_1,\ldots,p_m\}$ and $\{q_1,\ldots,q_{n-1}\}$ be bases of P and Q_0 respectively and assume that $P, Q_0 \subset C^{m+n-1}[0,\delta]$, $\delta > 0$. For each f, define

$$\phi_i \equiv \phi_i(f) = \begin{cases} p_i & \text{if} \quad 1 \leq i \leq m \\ fq_{i-m} & \text{if} \quad m+1 \leq i \leq m+n-1 \end{cases}$$

where $m \geq 1$ and $n \geq 2$. As a consequence of Theorem 4.3, we have the following result (cf. [35]).

THEOREM 4.5. Let $f \in C^{m+n-1}[0,\delta]$, where $\delta > 0$, be such that the Wronskian matrix at 0

$$(4.6) \quad B_{m+n-1}(f) = \begin{bmatrix} \phi_1(0) & \cdots & \phi_{m+n-1}(0) \\ \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot \\ \phi_1^{(m+n-2)}(0) & \cdots & \phi_{m+n-1}^{(m+n-2)}(0) \end{bmatrix}$$

is nonsingular. For each $\varepsilon > 0$, $0 < \varepsilon \leq \delta$, let $(p_\varepsilon, q_\varepsilon)$ be the (unique) minimizing pair for f from $P \times Q_1$ as described above. Then the net $\{(p_\varepsilon, q_\varepsilon)\}$ converges, as $\varepsilon \to 0^+$, to a pair $(p_0, q_0) \in P \times Q_1$. Furthermore,

$$\left(f - \frac{p_0}{q_0}\right)(x) = O(x^{m+n-1})$$

as $x \to 0^+$.

The uniqueness of the minimizing pairs follows, since $\det B_{m+n-1}(f) \neq 0$ implies that the space R_f is Haar in $[0, \varepsilon_0]$ for some $\varepsilon_0 > 0$. In [35], $(p_0, q_0) \equiv (p_0(f), q_0(f))$ is called the best local quasi-rational approximant of f from $P \times Q_1$. As an example, if we take P and Q to be the subspaces of polynomials of degree m and n respectively, the above theorem gives the following result which relates to $[m/n]$ Padé approximants.

COROLLARY 4.1. *Let* f *be in* $C^{m+n+1}[0,\delta]$ *for some* $\delta > 0$, *and write* f *as in* (4.1). *Suppose that the normality condition* (4.2) *of* f *is satisfied*. *Then there is an* ε_0 , $0 < \varepsilon_0 \leq \delta$, *such that for each* ε , $0 < \varepsilon \leq \varepsilon_0$, *there exists a unique pair of polynomials* $(p_\varepsilon, q_\varepsilon)$, *with* degree $p_\varepsilon \leq m$, degree $q_\varepsilon \leq n$ *and* $q_\varepsilon(0) = 1$, *such that* $(p_\varepsilon, q_\varepsilon)$ *solves the minimization problem* (4.5), *and such that the net* $\{(p_\varepsilon, q_\varepsilon)\}$ *converges*, *as* $\varepsilon \to 0^+$, *uniformly on some* $[0, \varepsilon_1]$, $\varepsilon_1 > 0$, *to a pair of polynomials* (p_0, q_0) , *with* degree $p_0 \leq m$, degree $q_0 \leq n$ *and* $q_0(0) = 1$. *Furthermore, the rational function* p_0/q_0 *is the* $[m/n]$ *Padé approximant of* f.

For more information on the problem of best local approximation an interested reader should consult [35, 36].

References

1 Allen, G.D., C.K. Chui, W.R. Madych, F.J. Narcowich and P.W. Smith, Padé approximation and orthogonal polynomials, Bull. Aust. Math. Soc., 10 (1974), 263-271.

2 Allen, G.D., C.K. Chui, W.R. Madych, F.J. Narcowich and P.W. Smith, Padé approximation and Gaussian quadrature, Bull. Aust. Math. Soc., 11 (1974), 63-71.

3 Allen, G.D., C.K. Chui, W.R. Madych, F.J. Narcowich and P.W. Smith, Padé approximation of Stieltjes series, J. Approx. Theory, 14 (1975), 302-316.

4 Allen, G.D. and F.J. Narcowich, On the representation and approximation of a class of operator-valued analytic functions, Bull. Amer. Math. Soc. 81 (1975), 410-413.

5 Arms, R.J. and A. Edrei, The Padé tables and continued fractions generated by totally positive sequences, Math. Essays dedicated to A.J. Macintyre, Ohio Univ. Press, Athens (Ohio), 1970.

6 Baker, G.A. Jr., Advances in Theoretical Physics, K.A. Brueckner, ed., Vol. 1, Academic Press, New York, 1965.

7 Baker, G.A. Jr., Essentials of Padé Approximants, Academic Press, London-New York, 1975.

8 Baker, G.A. Jr., Convergence of Padé approximants using the solution of linear functional equations, J. Math. Phys., 16 (1975), 813-822.

9 Baker, G.A. Jr., A theorem on the convergence of Padé approximants, Studies in Applied Math. (To appear).

10 Baker, G.A. Jr. and J.L. Gammel, ed., The Padé Approximant in Theoretical Physics, Academic Press, New York-London, 1970.

11 Baker, G.A. Jr. and P.R. Graves-Morris, Convergence of rows of the Padé table, J. Math. Analysis and Appls. (To appear).

12 Balk, M.B. The Padé interpolation process for certain analytic functions, Issledovaniya po souremennym problemam teorii funckcii kompleksnogo peremennogo, Gosudarstr. Izdat. Fiz.-Mat. Lit., Moscow, 1960.

13 Barnsley, M., The bounding properties of the multipoint Padé approximant to a series of Stieltjes on the real line, J. Math. Phys., 14 (1973), 299-313.

14 Basdevant, J.L., D. Bessis, and J. Zinn-Justin, Padé approximants in strong interactions. Two-body pion and kaon systems, Il Nuovo Cimento, Series X, 60 A (1969), 185-238.

15 Beardon, A.F., The convergence of Padé approximants, J. Math. Analysis and Applications, 21 (1968), 344-346.

16 Bessis, D., Topics in the theory of Padé approximants, in Padé Approximants, Graves-Morris, ed., The Institute of Physics, London, 1973.

17 Bessis, D., Padé approximants in Quantum Field Theory, in Padé Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

18 Bessis, D., Construction of variational bounds for the n-body eigenstate problem by the method of Padé approximations, in Applications of the Padé Method in Mechanics, Springer Verlag Lecture Notes in Physics, (To appear).

19 Bessis, J.D. and J.D. Talman, Variational approach to the theory of operator Padé approximants, Rocky Mountain J. Math., 4 (1974), 151-158.

20 Bessis, D. and M. Villani, Padé approximation in non-relativistic and relativistic quantum mechanics, International Conference on Mathematical Problems of Quantum Field Theory and Quantum Statistics, Moscow, 1972.

21 Bove, A., G. Fano, A.G. Teolis and G. Turchetti, Functional integration, Padé approximants, and the Schrödinger equation, J. Math. Phys., 16 (1975), 268-270.

22 Brezinski, C., Forme confluente de l'ε-algorithme topologique, Numer. Math. 23 (1975), 363-370.

23 Brezinski, C., Analyse numérique.—Génération de suites totalement monotones et oscillantes, C. R. Acad. Sc. Paris, t. 280, Series A, (1975), 729-730.

24 Brezinski, C., Rhombus algorithms connected to the Padé table and continued fractions, (To appear).

25 Brezinski, C., Computation of Padé approximants and continued fractions, (In manuscript).

26 Bruin, M.G. de, Generalized C-fractions and a multidimensional Padé table, Ph.D. Thesis, Amsterdam, 1974.

27 Bruin, M.G. de, Three classes of Padé tables whose upper halves are normal, (In manuscript).

28 Bruin, M.G. de, Convergence in the Padé table of ${}_1F_1(1;c;x)$ with $c \notin Z \setminus N$, (In manuscript).

29 Bruin, M.G. de and H. van Rossum, Formal Padé approximation, Nieuw Arch. voor Wiskunde (3), 23 (1975), 115-130.

30 Carleman, T., Les fonctions quasi-analytiques, Gauthier-Villars, Paris, 1926.

31 Chisholm, J.S.R., Rational approximants defined from double power series, Math. Comp., 27 (1973), 841-848.

32 Chisholm, J.S.R. and P.R. Graves-Morris, Generalizations of the theorem of de Montessus to two variable approximants, (To appear).

33 Chisholm, J.S.R. and J. McEwan, Rational approximants defined from power series in N variables, Proc. R. Soc. Lond. A, 336 (1974), 421-452.

34 Chui, C.K., O. Shisha and P.W. Smith, Padé approximants as limits of chebyshev rational approximants, J. Approx. Theory, 12 (1974), 201-204.

35 Chui, C.K., O. Shisha and P.W. Smith, Best local approximation, J. Approx. Theory, (To appear).

36 Chui, C.K., P.W. Smith and J.D. Ward, Best L_2 local approximation, (To appear).

37 Claessens, G., A new look at the Padé table and the different methods for computing its elements, J. Comp. and Appl. Math., 1 (1975), 141-152.

38 Claessens, G., The rational Hermite interpolation problem and some related recurrence formulas, (In manuscript).

39 Claessens, G., A new algorithm for osculatory rational interpolation, (In manuscript).

40 Clenshaw, C.W. and K. Lord, Rational approximations from Chebyshev series, Lanczos Festschrift, (To appear).

41 Cody, W.J., A survey of practical rational and polynomial approximation of functions, Studies in Applied Mathematics, 6 (1970), 86-109.

42 Common, A.K. and P.R. Graves-Morris, Some properties of Chisholm approximants, J. Inst. Maths. Applics., 13 (1974), 229-232.

43 Copley, L.A. and D. Masson, Padé-approximant calculation of π-π scattering, Phys. Review, 164 (1967), 2059-2062.

44 Corbella, O.D., C.R. Garibotti, and F.F. Grinstein, Rational polynomial approximants and scattering amplitude for long range potentials, (In manuscript).

45 Critchfield, C.L. and J.L. Gammel, Rational approximants for inverse functions of two variables, Rocky Mountain J. Math., 4 (1974), 339-349.

46 Dahlquist, G., S.C. Eisenstat and G.H. Goloub, Bounds for the error of linear systems of equations using the theory of moments, J. Math. Anal. and Appl., 34 (1972), 151-166.

47 Donnelly, J.D.P., The Padé table, in Methods of Numerical Approximation, D.C. Handscomb, ed., Pergamon Press, Oxford, 1966.

48 Ellis, H.G., Continued fraction solutions of the general Riccati differential equation, Rocky Mountain J. Math., 4 (1974), 353-356.

49 Edrei, A. The Padé table of meromorphic functions of small order with negative zeros and positive poles, Rocky Mountain J. Math. 4 (1974), 175-180.

50 Edrei, A., The Padé table of functions having a finite number of essential singularities, Pac. J. Math., 56 (1975), 429-453.

51 Edrei, A., Convergence of the complete Padé tables of trigonometric functions, J. Approx. Th., 15 (1975), (To appear).

52 Edrei,A., The complete Padé tables of certain series of simple fractions, Rocky Mountain J. Math., (To appear).

53 Fair, W., Padé approximation to the solution of the Ricatti equation, Math. of Comp., 18 (1964), 627-634.

54 Fair, W., and Y.L. Luke, Padé approximations to the operator exponential, Numer. Math., 14 (1970), 379-382.

55 Field, D.A., Series of Stieltjes, Padé approximants and continued fractions, (To appear).

56 Fisher, M.E., Critical point phenomena—The role of series expansions, Rocky Mountain J. Math., 4 (1974), 181-201.

57 Fisher, M.E. and H.B. Tarko, Tests of strong scaling in the three-dimensional Ising model, Physical Review B, 11 (1975), 1131-1133.

58 Fleischer, J., J.L. Gammel, and M.T. Menzel, Matrix Padé approximants for the $^1S_0$ and $^3P_0$ partial waves in nucleon-nucleon scattering, Physical Review D, 8 (1973), 1546-1552.

59 Fleischer, J. and J.A. Tjon, Bethe-Salpeter equation for J=0 nucleon-nucleon scattering with one-boson exchange, Nuclear Physics, B 84 (1975), 375-396.

60 Franzen,N.R., Some convergence results for Padé approximants, J. Approx. Theory, 6 (1972), 254-263.

61 Franzen, N.R., Convergence of Padé approximants for a certain class of meromorphic functions, J. Approx. Theory, 6 (1972), 264-271.

62 Freud, G., Padé approximation and Gaussian quadrature, Joint University of Texas—Texas A&M University Seminar on Approximation Theory, 1975.

63 Frobenius, G., Über Relationen zwischen den Näherungsbrüchen von Potenzreihen, J. für reine und angew. Math., 90 (1881), 1-17.

64 Gallucci, M.A. and W.B. Jones, Rational approximations corresponding to Newton series (Newton-Padé approximants), J. Approx. Theory, (To appear).

65 Gammel, J.L., Continuation of functions beyond natural boundaries, Rocky Mountain J. Math., 4 (1974), 203-206.

66 Gammel, J.L., Effect of random errors (noise) in the terms of a power series on the convergence of the Padé approximants, (To appear).

67 Gammel, J.L., Review of two recent generalizations of the Padé approximant, Padé Approximants and Their Applications, edited by P.R. Graves-Morris, Academic Press, 1973.

68 Gammel, J.L. and M.T. Menzel, Bethe-Salpeter equation for nucleon-nucleon scattering matrix Padé approximants, Physical Review D, 7 (1973), 565-568; II. Physical Review D, 11 (1975), 963-966.

69 Gammel, J.L. and J. Nuttall, Convergence of Padé approximants to quasianalytic functions beyond natural boundaries, J. Math. Analysis and Appl., 43 (1973), 694-696.

70 Gammel, J.L., C.C. Rousseau, and D.P. Saylor, A generalization of the Padé approximant, J. Math. Anal. and Appl., 20 (1967), 416-420.

71 Garibotti, C.R. and F.F. Grinstein, A summation procedure for expansions in orthogonal polynomials, (In manuscript).

72 Garibotti, C.R. and F.F. Grinstein, Summation of the Born series in potential scattering, (In manuscript).

73 Genz, A., The ε-algorithm and some other applications of Padé approximants in numerical analysis, in Padé Approximants and Their Applications, P.R. Graves-Morris, ed., Academic Press, London-New York, 1973.

74 Gilewicz, J., Method of the determination of the conformal mapping by the accelerated iterations with a Padé approximant, (unpublished manuscript).

75 Gilewicz, J., Totally monotonic and totally positive sequences for the Padé approximants method, CPT-CNRS Marseille, (To appear).

76 Gilewicz, J., Meilleur approximant de Padé en analyse numérique, (To appear).

77 Gilewicz, J., Numerical detection of the best Padé approximant and determination of the Fourier coefficients of the insufficiently sampled functions, in Padé Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

78 Golden, J.E., J.H. McGuire, and J. Nuttall, Calculating Bessel functions with Padé approximants, J. Math. Anal. and Appl., 43 (1973), 754-767.

79 Gončar, G.A., (Unpublished manuscript).

80 Gragg, W.B., Truncation error bounds for π-fractions, Bull. Amer. Math. Soc., 76 (1970), 1091-1094.

81 Gragg, W.B., The Padé table and its relation to certain algorithms of numerical analysis, SIAM Review, 14 (1972), 1-62.

82 Gragg, W.B., Matrix interpretations and applications of the continued fraction algorithm, Rocky Mountain J. Math., 4 (1974), 213-225.

83 Gragg, W.B. and G. D. Johnson, The Laurent-Padé table, Information Process, 74 (1974), 632-637.

84 Graves-Morris, P.R. (ed.), Padé Approximants and Their Applications, Proc. Conference at University of Kent 1972, Academic Press, London-New York, 1973.

85 Graves-Morris, P.R., R. Hughes Jones and G.J. Makinson, The calculation of some rational approximants in two variables, J. Inst. Maths. Applics, 13 (1974), 311-320.

86 Graves-Morris, P.R. and D.E. Roberts, Calculation of Canterbury approximants, (In manuscript).

87 Heine, E., Handbuch der Kugelfunctionen: Erster Band, Theorie der Kugelfunctionen und der verwandten Functionen; Zweiter Band, Anwendunger der Kugelfunctionen und der verwandten Functionen, G. Reimer, Berlin, 1878, 1881.

88 Henrici, Peter and Pia Pfluger, Truncation error estimates for Stieltjes fractions, Num. Mathematik, 9 (1966), 120-138.

89 Holdaman, J.T. Jr., A method for the approximation of functions defined by formal series expansions in orthogonal polynomials, Math. Comp. 23 (1969), 275-288.

90 Jager, H., A multidimensional generalization of the Padé table, Ph.D. Thesis, Amsterdam, 1964.

91 Jones, R. H., General rational approximants in n variables, (In manuscript).

92 Jones, W.B. and W.J. Thron, eds., Proceedings of the International Conference on Padé Approximants, Continued Fractions and Related Topics, Rocky Mountain J. Math., 4 (1974), no. 2.

93 Jones, W.B. and W.J. Thron, On convergence of Padé approximants SIAM J. Math. Analysis, 6 (1975), 9-16.

94 Karlsson, J., Rational interpolation with free poles in the complex plane, Preprint University of Umeå, Department of Mathematics, no. 6, 1972.

95 Karlsson, J., Rational interpolation and best rational approximation, Preprint University of Umeå, Department of Mathematics, no. 1, 1974.

96 Karlsson, J., Rational Hermite interpolation procedures, Preprint University of Umea, Department of Mathematics, no. 2, 1974.

97 Karlsson, J. and B. von Sydow, The convergence of Padé approximants to series of Stieltjes, (In manuscript).

98 Knowles, J.B. and D.W. Leggett, An analysis of transport delay simulation methods, Radio and Electronic Engineer, 42 (1972), 172-178.

99 Longman, I.M., Computation of the Padé table, Intern. J. Computer Math. Section B, 3 (1971), 53-64.

100 Longman, I.M., On the generation of rational function approximations for Laplace transform inversion with an application to viscoelasticity, SIAM J. Appl. Math., 24 (1973), 429-440.

101 Longman, I.M., Best rational function approximation for Laplace transform inversion, SIAM J. Math. Anal., 5 (1974), 574-580.

102 López, C. and F.J. Ynduráin, Model independent calculation of K_p dispersion relations: Extrapolation with Padé techniques, Nuclear Physics, B 64 (1973), 315-333.

103 López, C. and F.J. Ynduráin, The moment problem and stable extrapolations with an application to forward K_p dispersion relations, in Padé Approximants and Their Appl., Graves-Morris, ed., Academic Press, New York, 1974.

104 Luke, Y.L., The Padé table and the τ-method, J. Math. and Phys., 37 (1958), 110-127.

105 Luke, Y.L., The Special Functions and Their Applications, vols. 1 and 2, Academic Press, New York, 1969.

106 Luke, Y.L., Mathematical Functions and Their Approximations, Academic Press, New York, 1975.

107 Luke, Y.L., On the error in the Padé approximants for a form of the incomplete Gamma function including the exponential function, SIAM J. Math. Anal. (1975), 829-839.

108 Luke, Y.L., W. Fair and J. Wimp, Predictor-Corrector formulas based on rational interpolants, Comp. & Maths. with Appls., 1 (1975), 3-12.

109 Maehly, H., First Interim Progress Report on Rational Approximations, Project NR. 044-196, Princeton University, June 23, 1958.

110 Magnus, A., Certain continued fractions associated with the Padé table, Math. Zeitschr., 78 (1962), 361-374.

111 Magnus, A., P-fractions and the Padé table, Rocky Mountain J. Math., 4 (1974), 257-259.

112 Magnus, A. and J. Wynn, On the Padé table of cos z, Proc. Amer. Math. Soc., 47 (1975), 361-367.

113 Markoff, A.A. Differenzenrechnung, B.G. Teubner, Leipzig, 1896.

114 Masson, D., Padé approximant and partial-wave integral equation, J. Math. Phys., 8 (1967), 512-514.

115 Masson, D., Hilbert space and the Padé approximant, in The Padé Approximant in Theoretical Physics, Baker Jr. and Gammel, eds., Academic Press, New York, 1970.

116 Masson, D., Padé approximants and Hilbert spaces, in Padé Approximants and Their Applications, Graves-Morris, ed., Academic Press, New York, 1973.

117 McGuire, J.H., J.E. Golden and J Nuttall, Solution of the schrödinger equation using Padé approximants to sum asymptotic series, Rocky Mountain J. Math., 4 (1974), 375-376.

118 Milošević, S. and H.E. Stanley, Calculation of the equation of state near the critical point for the Heisenberg model using Padé approximants, in Padé Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

119 Montessus de Ballore, R. de, Sur les fractions continues algébriques, Bull. Soc. Math. France, 30 (1902), 28-36.

120 Narcowich, F.J. and G.D. Allen, R-Operators I. Representation theory and applications, Indiana J. Math., (To appear).

121 Nuttall, J., The convergence of Padé approximants of meromorphic functions, J. Math. Analysis and Appl., 31 (1970), 147-153.

122 Nuttall, J., The connection of Padé approximants with stationary variational principles and the convergence of certain Padé approximants, in The Padé Approximants in Theoretical Physics, G.A. Baker Jr., ed., Academic Press, New York, 1970.

123 Nuttall, J., Variational principles and Padé approximants, in Padé Approximants and Their Applications, Graves-Morris, ed., Academic Press, New York, 1973.

124 Nuttall, J., The convergence of certain Padé approximants, Rocky Mountain J. Math., 4 (1974), 269-272.

125 Nuttall, J., The convergence of Padé approximants for a class of functions with branch points, (In manuscript).

126 Nuttall, J., Orthogonal polynomials for complex weight functions and the convergence of related Padé approximants, (In manuscript).

127 Padé, H., Sur la représentation approchée d'une fonction par des fractions rationelles, Ann. Sci., Ecole Norm. Sup. (3), 9 (1892), Supplément 1-93.

128 Padé, H., Sur les series entières convergentes ou divergentes, Acta Math. 18 (1894), 97-111.

129 Padé, H., Sur la généralisation des fractions continues algebrique, J. de Math. Pures et Appl., 10 (1894), 291-329.

130 Padé, H., Sur la distribution des réduites anormales d'une fonction, C.R. Acad. Sci. Paris, 130 (1900), 102-104.

131 Perron, O., Die Lehre von den Kettenbrüchen, Band II, B.G. Teubner, Stuttgart, 1957.

132 Pommerenke, C., Padé approximants and convergence in capacity, J. Math. Analysis and Appl., 41 (1973), 775-780.

133 Roberts, D., H.P. Griffiths and D.W. Wood, The analysis of double power series using Canterbury approximants, J. Phys. A : Math. Gen., 8 (1975), 1365-1372.

134 Ronveaux, A., Padé approximant and homographic transformation of Riccati Phase's equations, in Padé Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

135 Ronveaux, A., M.C. Dumont-Lepage, and J. Habay, Counting the poles of the solutions of a general Riccati equation, (In manuscript).

136 Rossom, H. van, Systems of orthogonal and quasi orthogonal polynomials connected with the Padé table, I, II, III, Nederl. Akad. Wetensch. Proc., 58 (1955), 517-525, 526-534, 676-682.

137 Rossum, H. van, On the poles of Padé approximations to e^z, Nieuw Archief voor Wiskunde, 19 (1971), 37-45.

138 Saff, E.B., An extension of Montessus de Ballore's theorem on the convergence of interpolating rational functions, J. Approx. Theory, 6 (1972), 63-67.

139 Saff, E.B. and R.S. Varga, Convergence of Padé approximants to e^{-z} on unbounded sets, J. Approx. Theory, 13 (1975), 470-488.

140 Saff, E.B. and R.S. Varga, On the zeros and poles of Padé approximants to e^z, Num. Math. (To appear).

141 Saff, E.B. and R.S. Varga, On the sharpness of theorems concerning zero-free regions for certain sequences of polynomials, Num. Math. (To appear).

142 Saff, E.B., R.S. Varga and W.-C. Ni, Geometric convergence of rational approximations to e^{-z} in infinite sectors, Num. Math. (To appear).

143 Salzer, H.E., Note on osculatory rational interpolation, Math. Comp., 16 (1962), 486-491.

144 Shafer, R.E., On quadratic approximation, SIAM J. Numer. Anal., 11 (1974), 447-460.

145 Shamash, Y., Linear system reduction using Padé approximation to allow retention of dominant modes, Int. J. Control, 21 (1975), 257-272.

146 Sobhy, M.I., Applications of Padé approximants in electrical network problems, in Pade Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

147 Stahl, H., Ph.D. Thesis, Technical University Berlin, (To be completed).

148 Stieltjes, T.-J., Recherches sur les fractions continues, Ann. Fac. Sci., Toulouse 8 (1894), 1-122 and 9 (1895), 1-47.

149 Szegö, G., Orthogonal polynomials, Amer. Math. Soc. Colloquium Publications, 23, Amer. Math. Soc., New York, 1939; revised ed., 1959.

150 Tarko, H.B. and M.E. Fisher, Theory of critical point scattering and correlations. III. The Ising model below T_c and in a field, Physical Review B, 11 (1975), 1217-1253.

151 Tjon, J.A., Padé approximants in three-body calculations, Physical Review D, 1 (1970), 2109-2112.

152 Tjon, J.A., Application of Padé approximants in the three-body problem, in Pade Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

153 Turchetti, G., Matrix Padé approximants in πN scattering, Lettere al Nuovo Cimento, 6 (1973), 497-503.

154 Underhill, C., and A. Wragg, Convergence properties of Padé approximants to exp(z) and their derivatives, J. Inst. Math. Appl., 11 (1973), 361-367.

155 Waadeland, H., T-fractions from a different point of view, Rocky Mountain J. Math., 4 (1974), 391-393.

156 Waadeland, H., Two convergence theorems for T-fractions, (Unfinished manuscript).

157 Wall, H.S., The Analytic Theory of Continued Fractions, D. van Nostrand, Toronto-New York-London, 1948.

158 Wallin, H., On the convergence theory of Padé approximants, ISNM, 20 (1972), 461-469.

159 Wallin, H., The convergence of Padé approximants and the size of the power series coefficients, Applicable Analysis, 4 (1974), 235-251.

160 Walsh, J.L., On approximation to an analytic function by rational functions of best approximation, Math. Z., 38 (1934), 163-176.

161 Walsh, J.L., Padé approximants as limits of rational functions of best approximation, J. Math. Mech. 13 (1964), 305-312.

162 Walsh, J.L., Padé approximants as limits of rational functions of best approximation, real domain, J. Approx. Theory, 11 (1974), 225-230.

163 Walsh, J.L., The role of the pole in rational approximation, Rocky Mountain Math. J., 4 (1974), 283-286.

164 Warner, D.D., Hermite interpolation with rational functions, Ph.D. Thesis, University of California at San Diego, 1974.

165 Warner, D.D., An extension of Saff's theorem on the convergence of interpolating rational functions, (To appear).

166 Wheeler, J.C., Modified moments and Gaussian quadratures, Rocky Mountain J. Math., 4 (1974), 287-296.

167 Wheeler, J.C., M.G. Prais, and C. Blumstein, Analysis of spectral densities using modified moments, Physical Review B, 10 (1974), 2429-2447.

168 Wilson, R., Divergent continued fractions and non-polar singularities, Proc. London Math. Soc., 30 (1930), 38-57.

169 Wuytack, L., The use of Padé approximation in numerical integration, Proc. Colloquium Euromech 58, on the Method of Padé and its applications to Mechanics, Toulon, 1975.

170 Wuytack, L., On the osculatory rational interpolation problem, Math. Comp., 29 (1975), 837-843.

171 Wuytack, L., Eigenschaften eines Algorithmus zur rationalen Interpolation, ISNM 26 (1975), 193-199.

172 Wuytack, L., Numerical integration by using nonlinear techniques, J. Comp. and Appl. Math., (To appear).

173 Wynn, P., L' ε-algorithms e la tavola di Padé, Rend. di Mat. Roma, 20 (1961), 403-408.

174 Wynn, P., Upon systems of recursions which obtain among the quotients of the Padé table, Num. Math., 8 (1966), 264-269.

175 Wynn, P., Some recent developments in the theories of continued fractions and the Padé table, Rocky Mountain J. Math., 4 (1974), 297-324.

176 Yndurain, F.J., Stable extrapolation of scattering amplitudes using unitarity, Annals of Physics, 75 (1975), 171-218.

177 Yndurâin, F.J., Convergence of regularized Padé approximants for semi-infinite range Stieltjes function, Presented to the Euromech Colloquium, Toulon (France).

178 Zakian, V., Properties of I_{MN} approximants, in Padé Approx. and Their Appl., Graves-Morris, ed., Academic Press, New York, 1973.

179 Zakian, V., Properties of I_{MN} and J_{MN} approximants and applications to numerical inversion of Laplace transforms and initial value problems, J. Math. Anal. and Appl., 50 (1975), 191-222.

180 Zinn-Justin, J., Convergence of Padé approximants in the general case, Rocky Mountain J. Math., 4 (1974), 325-330.

Supplementary references

1 Atkinson, F.V., On lacunary and other orthogonal polynomials, Math. y Fisica Teorica, 10 (1954), 97-110.

2 Atkinson, F.V., Orthogonal polynomials and lacunary approximants, in Padé Approximants and Their Appl., Graves-Morris, ed., Academic Press, London and New York, 1973.

3 Chisholm, J.S.R. and R. H. Jones, Relative scale co-variance of N-variable approximants, Proc. R. Soc. Lond. A. 344 (1975), 465-470.

4 Engels, J. and J. Fleischer, Legendre Padé approximants in πN Scattering, Il Nuovo Cimento, 24 (1974), 73-92.

5 Field, D.A., Nested sequences of lens-shaped regions which contained Padé and continued fraction approximations to a series of Stieltjes, (Submitted for publication).

6 Fleischer, J., Analytic continuation of scattering amplitudes and Padé approximants, Nuclear Phy. B 37 (1972), 59-76.

7 Fleischer, J., Generalizations of Padé approximants, in Padé Approximants, Graves-Morris, ed., The Institute of Physics, London, 1973.

8 Fleischer, J., Nonlinear Padé approximants for Legendre series, J. Math. Phys., 14 (1973), 246-248; and in Padé Approximants and Their Appls., Graves-Morris, ed., Academic Press, London-New York, 1973.

9 Graves-Morris, P.R. and R.H. Jones, An analysis of two variable rational approximants, (In manuscript).

10 Nuttall, J. and S.R. Singh, Extension of proof of convergence of Padé approximants for a class of functions with branch points, (In manuscript).

11 Watson, P.J.S., Two-variable rational approximants: a new method, J. Phys. A : Nucl. Gen., 7 (1974), L167-L170.

Charles K. Chui*
Department of Mathematics
Texas A&M University
College Station,
Texas, 77840

*Supported in part by the U. S. Army Research Office under Grant Number DAHC04-75-G-0186.

DEGREE OF APPROXIMATION

Ronald A. DeVore

We examine techniques for estimating the degree of approximation of a function in terms of its smoothness. This is done in the settings of polynomial and spline approximation in L_p, pointwise and local estimates in C, as well as some remarks on constrained approximation.

1 Introduction

A major portion of the theory of approximation of functions is concerned with the connections between the structural properties of a function and its degree of approximation. The objective is to relate the smoothness of the function to the rate of decrease of the degree of approximation to zero. We are interested here in examining these questions for polynomial approximation, both algebraic and trigonometric, as well as spline approximation. These are then the most classical settings where the results are the most penetrating and satisfying.

We will limit ourselves to only one aspect of the problem, namely, the derivation of estimates for the degree of approximation in terms of smoothness. These are the direct estimates of approximation which are also commonly called Jackson Theorems, after D. Jackson, who first obtained such results for trigonometric approximation. We will measure smoothness using the r-th order modulus of smoothness. Estimates given for the r-th order modulus of smoothness automatically include lower order estimates in terms of the j-th order modulus of smoothness of $f^{(k)}$, with $j+k\leq r$.

Our desire here is not to try to be encyclopedic or sur-

Preceding page blank

vey the subject, but rather we emphasize techniques which are simple, general, and give sharp results. We strike out then to give, as much as possible, a unified viewpoint to obtaining direct theorems of approximation. Wherever possible, we seek linear methods to do the job.

Perhaps the most useful technique in deriving direct estimates is smoothing, where the arbitrary function f is approximated first by a smooth function and then the smooth function is approximated by the appropriate polynomials or splines. In such considerations, the actual method of smoothing is not important, although sometimes it is desirable to do the smoothing with linear operators. What is important is the connection between the error in approximation and the deterioration of the smoothness. For example, if we approximate f by a function g with $g^{(r)}$ in L_p, then we are interested in how small the norm of f-g can be made compared to the norm of $g^{(r)}$. One way of making such comparisons is through the Peetre K-functional as described in Section 2.

There is a very intimate connection between the Peetre K-functional and the more classical modulus of smoothness as is put forth in Theorem 2.1. From this theorem, it follows that for most problems in the degree of approximation, the K-functional and the modulus of smoothness are the same. Hence, in seeking estimates for the degree of approximation, we can just as well obtain estimates in terms of the K-functional. This will mean that it suffices to prove estimates for smooth functions, which can significantly simplify matters.

In Section 3, we discuss local approximation by polynomials of fixed degree. The main theme is to examine the dependence of the degree of approximation on the length of the interval. This is easy enough to obtain, but nevertheless important as building blocks for spline approximation. Moreover,

in some sense, even polynomial approximation can be viewed as piecing together local approximations at the expense of letting the degree of the polynomials grow.

We discuss spline approximation in Section 4. There are several ways to generate good spline approximations. We restrict our attention to two of them given by V. Popov - Bl. Sendov and C. deBoor - G. Fix respectively. These approaches expose readily the highly local nature of spline approximation.

Trigonometric polynomial approximation is discussed briefly in Section 5. Here, the results are well known and easy to obtain. The coresponding case of algebraic polynomial approximation requires more care and effort. The Jackson estimates for the space $C[-1,1]$ can be gotten simply by reverting back to the periodic case via the usual substitution $x=\cos\theta$. We however give a direct construction in the algebraic case since this carries over to general L_p. This is done in Section 6.

A satisfactory description of algebraic polynomial approximation requires the use of pointwise estimates since the approximation can be improved near the end points of the interval. We look at such pointwise estimates in Section 7 for $C[-1,1]$. Here, we find it useful to revert back to the trigonometric case since all the constructions can be made with modifications of the classical Jackson operators.

In Section 8, we look at the constrained problem of monotone approximation. While the addition of the constraint makes the problem much more difficult, some of the general lines of attack still carry over.

A word about proofs. In all instances at least a sketch of the proof is given -- the most sketchy being for monotone approximation. In the absence of complete details, there is of course an appropriate reference to the literature. We use for notational convenience the convention that C always denotes a constant although not always the same one even in a given line.

The exact dependence of C on other parameters is always explicitly given.

2 Moduli of smoothness and the K-functional

Let $I=[-1,1]$ and for $1\le p\le\infty$, let $L_p(I)$ denote the set of real valued p-th power integrable functions on I, with $\|\cdot\|_p(I)$ the usual L_p norm on I. Similarily, let $L_p(2\pi)$ denote the set of 2π periodic real valued functions which are p-th power integrable on $[-\pi,\pi]$ and $\|\cdot\|_p(2\pi)$ the L_p norm on $[-\pi,\pi]$. The spaces $C(I)$ and $C(2\pi)$ are defined likewise with the L_∞ norm. We use the notation $\|\cdot\|_p$ without the I or 2π designation when there is no possibility of confusion.

If $f\varepsilon L_p(2\pi)$, $h>0$, and r is a positive integer, we define the r-th order modulus of smoothness of f as

$$(2.1)\quad \omega_{r,p}(f,h) = \sup_{o<t\le h}\|\Delta_t^r(f,x)\|_p,$$

with $\Delta_t^r(f)$ the r-th forward difference of f with step size t. Throughout, we keep the convention that norms are taken with respect to the variable x. For $f\varepsilon C(2\pi)$, we have

$$(2.2)\quad \omega_r(f,h) = \omega_{r,\infty}(f,h) = \sup_{o<t\le h}\|\Delta_t^r(f,x)\|_\infty.$$

In $L_p(I)$, the modulus of smoothness needs a little more care in its definition. When $t>0$, let $I_t=[-1,1-t]$. Then, for $f\varepsilon L_p(I)$, define

$$(2.3)\quad \omega_{r,p}(f,h) = \sup_{o<t\le h}\|\Delta_t^r(f,x)\|_p(I_{rt}).$$

Here, $\|\cdot\|_p(I_{rt})$ indicates that the L_p norm is taken over the interval I_{rt}. This notational convention will be used throughout the paper. For $f\varepsilon C(I)$, we have

$$(2.4)\quad \omega_r(f,h) = \omega_{r,\infty}(f,h) = \sup_{o<t\le h}\|\Delta_t^r(f,x)\|_\infty(I_{rt}).$$

The fundemental role of higher order moduli of smoothness in accurately describing degree of approximation was first recognized by A. Zygmund [36] in his characterization of functions approximated with order $O(n^{-1})$ as the set of functions with $\omega_2(f,t) = O(t)$. Actually, a modulus of smoothness can already in its very definition be considered as a measure of the degree of approximation of f by certain combinations of its translates. This of course is not a very interesting approximation process because the approximants are no simpler or smoother than the original function. However, the modulus of smoothness is also fundementally connected with approximation by smooth functions, as can be brought out best with the aid of the Peetre K-functional.

If r is a positive integer and $1 \leq p \leq \infty$, let $W_{r,p}(I)$ be the set of those functions f, with $f^{(r-1)}$ absolutely continuous and $f^{(r)}$ in $L_p(I)$. $W_{r,p}(2\pi)$ is defined similarily. For $f \varepsilon L_p(I)$, $1 \leq p \leq \infty$, define

$$(2.5)\quad K_{r,p}(f,h) = \inf_{g \varepsilon W_{r,p}(I)} \{\|f-g\|_p + h\|g^{(r)}\|_p\}.$$

For $f \varepsilon C(I)$, we have

$$(2.6)\quad K_r(f,h) = K_{r,\infty}(f,h) = \inf_{g \varepsilon W_{r,\infty}(I)} \{\|f-g\|_\infty + h\|g^{(r)}\|_\infty\}.$$

For a periodic function f, the K-functional $K_{r,p}(f,h)$ is defined similarily. The K-functional measures how well f can be approximated by smooth functions g with a control on the size of $\|g^{(r)}\|_p$.

For most purposes, the K-functional and the modulus of smoothness are equivalent for measuring the degree of approximation. Indeed, we have the following theorem of J. Peetre [27], H. Johnen [21] in the non-periodic case, (see also G. Freud [17], and G. Freud - V. Popov [19] for the non-per-

iodic case and $p=\infty$).

THEOREM 2.1. <u>There are constants</u> $C_1, C_2 > 0$ <u>depending only on</u> r, <u>such that if</u> $f \in L_p(I)$ (<u>or</u> $f \in L_p(2\pi)$), $1 \le p < \infty$, <u>and</u> $f \in C(I)$ (<u>or</u> $f \in C(2\pi)$), $p=\infty$, <u>then</u>

$$(2.7) \quad C_1 \omega_{r,p}(f,t) \le K_{r,p}(f,t^r) \le C_2 \omega_{r,p}(f,t), \qquad t>0.$$

<u>Proof</u>. The lower estimate in (2.7) is shown by using standard estimates for r-th differences. Let M_r denote the B-spline of order r (degree r-1) with knots at $0,1,\dots,r$ (see Section 4). Then, M_r is non-negative, vanishes outside of $(0,r)$, and we take a normalization so that $\int_0^r M_r(t)dt = 1$. M_r is the Peano kernel for r-th differences (see [9]) at the points $0,1,\dots,r$. Hence, if $g \in W_{r,p}(I)$,

$$(2.8) \quad \Delta_t^r(g,x) = t^r \int_{-\infty}^{\infty} g^{(r)}(x+u) M_r(t^{-1}u)t^{-1}du.$$

Using Minkowski's inequality for integrals and the fact that M_r has integral one, we can estimate for any $g \in W_{r,p}(I)$

$$\|\Delta_t^r(g,x)\|_p(I_{rt}) \le t^r \|g^{(r)}\|_p.$$

When $f \in L_p(I)$, $t>0$, and g is an arbitrary function from $W_{r,p}(I)$, then

$$\begin{aligned}\|\Delta_t^r(f,x)\|_p(I_{rt}) &\le \|\Delta_t^r(f-g,x)\|_p(I_{rt}) + \|\Delta_t^r(g,x)\|_p(I_{rt}) \\ &\le 2^r\{\|f-g\|_p + t^r \|g^{(r)}\|_p\}.\end{aligned}$$

Taking an infimum over all g on the right hand side and a supremum over all $0<t\le h$, we find

$$\omega_{r,p}(f,h) \le 2^r K_{r,p}(f,h^r).$$

The same reasoning goes over in the periodic case. Thus, we have proved the lower estimate in (2.7).

The other half of the inequality (2.7) requires the const-

ruction of a smooth approximation to f. This is simplest to do in the periodic case. To see the idea, consider first $r=1$. We can use Stecklov averages. Fix $t>0$ and for $f\in L_p(2\pi)$ define

$$g(x) = t^{-1} \int_0^t f(x+u)du.$$

Then,

$$\|f-g\|_p \le t^{-1} \int_0^t \|f(x+u)-f(x)\|_p du \le \omega_{1,p}(f,t).$$

Also, if F is a primitive of f, then $g(x) = t^{-1}(F(x+t)-F(x))$, and so $g'(x) = t^{-1}(f(x+t)-f(x))$. Hence,

$$t\|g'\|_p = \|f(x+t)-f(x)\|_p \le \omega_{1,p}(f,t).$$

Putting these last two inequalities together shows that

$$K_{1,p}(f,t) \le \|f-g\|_p + t\,\|g'\|_p \le 2\omega_{1,p}(f,t), \quad t>0,\ f\in L_p(2\pi).$$

For the general case of r, we use a type of higher averages. Let $f\in L_p(2\pi)$ and $t>0$, and now define

$$(2.9)\quad g(x) = \int_{-\infty}^{\infty} \{(-1)^{r+1}\Delta_u^r(f,x)+f(x)\}M_r(t^{-1}u)t^{-1}du$$

$$= f(x) + (-1)^{r+1} \int_{-\infty}^{\infty} \Delta_u^r(f,x)M_r(t^{-1}u)t^{-1}du.$$

Recalling that M_r is supported on $(0,r)$ and has integral one, we find that

$$(2.10)\quad \|f-g\|_p \le \int_{-\infty}^{\infty} \|\Delta_u^r(f,x)\|_p M_r(t^{-1}u)t^{-1}du$$

$$\le \int_0^r \|\Delta_{tu}^r(f,x)\|_p du \le \omega_{r,p}(f,rt).$$

Here, we changed variables to return to $(0,r)$ and used the fact that $M_r(u) \le 1$, for all u, in estimating the second inequality.

To estimate the r-th derivative of g, let F_1 be a primitive of f and in general F_k a primitive of F_{k-1} for $k \geq 2$. Then for a typical term in g, we have

$$\int_{-\infty}^{\infty} f(x+ku)M_r(t^{-1}u)t^{-1}du = \int_{-\infty}^{\infty} f(x+u)M_r\left((kt)^{-1}u\right)(kt)^{-1}du$$

$$= t^{-r}\Delta_{kt}(F_r,x),$$

because of (2.8). It follows that the $L_p(2\pi)$ norm of the derivative of such a term can be estimated by $t^{-r}\|\Delta_{kt}^r(f,x)\|_p$. This gives the following estimate for $g^{(r)}$.

$$(2.11) \quad t^r\|g^{(r)}\|_p \leq (2^r-1)\max_{1\leq k\leq r}\|\Delta_{kt}^r(f,x)\|_p \leq (2^r-1)\omega_{r,p}(f,rt).$$

The inequalities (2.10) and (2.11) combine to show that

$$K_{r,p}(f,t^r) \leq \|f-g\|_p + t^r\|g^{(r)}\|_p \leq 2^r\omega_{r,p}(f,rt) \leq C\omega_{r,p}(f,t)$$

with C a constant depending only on r. This is the right hand side of (2.7) for periodic functions.

The right hand side of (2.7) for I is a little more difficult to obtain due to the lack of periodicity. There are several possible approaches. One way is to extend the function f to a larger interval and then use the same approach as we have just given for the periodic case. This was done by G. Freud - V-Popov [19], for the case that $p=\infty$, with an extension that depends on t. It is also possible following H. Whitney [33] and O. Besov [3] to extend f to a larger interval by using a linear operator T and retaining the smoothness of f. This extension will be discussed at the end of this section.

It was recently pointed out to me by H. Johnen and K. Scherer how a simple modification of the technique given in the periodic case, already handles the non-periodic case as well. This avoids completely the problem of extending f. In order to point out this proof, let us first record what can be shown

using exactly the same reasoning as given above for the periodic case.

LEMMA 2.1. *If* $f \in L_p(I)$, $t>0$, *and* $[a,b+r^2t] \subseteq I$, *then with* $J=[a,b]$, *there is a function* $g \in W_{r,p}(J)$ *such that*

$$(2.12)\quad \|f-g\|_p(J) \leq \int_0^r \|\Delta_{tu}^r(f,x)\|_p(J)du \leq C\omega_{r,p}(f,t),$$

$$(2.13)\quad t^r\|g^{(r)}\|_p(J) \leq (2^r-1) \max_{1\leq k\leq r} \|\Delta_{kt}^r(f,x)\|_p(J) \leq C\omega_{r,p}(f,t),$$

with C *a constant depending only on* r. *Similarly, if* $[a-r^2t,b] \leq I$, *then there is a function* g *so that* (2.12) *and* (2.13) *hold with* Δ_{tu}^r *replaced by* Δ_{-tu}^r *in* (2.12) *and* Δ_{kt}^r *replaced by* Δ_{-kt}^r *in* (2.13).

Proof. The proof is exactly the same as in the periodic case with (2.12) coresponding to (2.10) and (2.13) coresponding to (2,11). The function g is defined as in (2.9) when $[a,b+r^2t] \subseteq I$ and as in (2.9) except that Δ_{-u}^r is used in place of Δ_u^r in the case that $[a-r^2t,b] \subseteq I$.

To obtain the estimate for the K-functional from (2.12) and (2.13), let ϕ be a function with $\phi(x)=0$, $-1\leq x\leq -\frac{1}{2}$, $\phi(x)=1$, $\frac{1}{2}\leq x\leq 1$, $0\leq\phi(x)\leq 1$, $x\in I$, and $\|\phi^{(i)}\|_\infty \leq C$, $i=0,1,\ldots,r$, with C depending only on r. Let $2r^2t<1$, g_1 be the function guaranteed by Lemma 2.1 for the interval $J_1=|-1,-\frac{1}{2}|$ and g_2 the coresponding function for the interval $J_2=|-\frac{1}{2},1|$. The function $g = (1-\phi)g_1 + \phi g_2 = g_1 + \phi(g_2-g_1)$ satisfies

$$(2.14)\quad \|f-g\|_p \leq \|f-g_1\|_p(J_1) + \|f-g_2\|_p(J_2) \leq C\omega_{r,p}(f,t),$$

because of (2.12).

We need also an estimate for $g^{(r)}$. The crucial interval is $J_3=[-\frac{1}{2},\frac{1}{2}]$, where we have

$$\|g^{(r)}\|_p(J_3) \leq C\{\|g_1^{(r)}\|_p(J_1) + \max_{0\leq i\leq r} \|g_1^{(i)}-g_2^{(i)}\|_p(J_3)\}$$

$$\le C\{\|g_1^{(r)}\|_p(J_1) + \|g_2^{(r)}\|_p(J_2) + \|g_1-g_2\|_p(J_3)\}$$
$$\le C\{\|g_1^{(r)}\|_p(J_1) + \|g_2^{(r)}\|_p(J_2) + \max_{i=1,2}\|f-g_i\|_p(J_3)\}.$$

Here, we have used the well known fact that for any h defined on J_3, we have $\|h^{(i)}\|_p(J_3) \le C\{\|h\|_p(J_3) + \|h^{(r)}\|_p(J_3)\}$, $1 \le i \le r$ with C depending only on r.

Our inequality for $g^{(r)}$ clearly also holds on $[-1,-\frac{1}{2}]$ and $[\frac{1}{2},1]$. Hence,

$$(2.15)\quad t^r\|g^{(r)}\|_p \le C \max_i \{\|f-g_i\|_p(J_i) + \|g_i^{(r)}\|_p(J_i)\}$$
$$\le C\omega_{r,p}(f,t),$$

where we have used (2.12) and (2.13). The inequalities (2.14) and (2.15) show that for r^2t

$$K_{r,p}(f,t^r) \le \|f-g\|_p + t^r\|g^{(r)}\|_p \le C\omega_{r,p}(f,t).$$

This inequality automatically also holds for $2r^2t<1$ because $K_{r,p}$ is sub-additive and $\omega_{r,p}$ is monotone. Thus we have proved (2.7) in the non-periodic case as well.

In our above proof, we could actually have given a finer estimate in that we could have used the middle estimates from (2.12) and (2.13). This gives the following lemma.

LEMMA 2.2. <u>If</u> $f\varepsilon L_p(I)$, $0\le t\le\frac{1}{4}r^{-2}$, <u>then there is a</u> $g\varepsilon W_{r,p}(I)$ with

$$(2.16)\quad \|f-g\|_p \le C\int_0^r \|\Delta_{tu}^r(f,x)\|_p(I_{rtu})du$$

$$(2.17)\quad t^r\|g^{(r)}\|_p \le C \max_{0\le k\le r}\|\Delta_{kt}^r(f,x)\|_p(I_{rkt})$$
$$+ \int_0^r \|\Delta_{tu}^r(f,x)\|_p(I_{rtu})du\ ,$$

<u>with</u> C <u>depending only on</u> r.

Proof. Let $g = (1-\phi)g_1 + \phi g_2$ as above and estimate $\|f-g\|_p$ as in (2.14) except now use the middle term in (2.12). Note that $\Delta_{-s}^r(f,x) = (-1)^r\Delta_s^r(f,x-rs)$. To estimate $\|g^{(r)}\|_p$, we argue exactly as in (2.15) except that we now use the middle terms of (2.12) and (2.13).

Many of the approximation processes developed in the following sections will show how polynomials and splines can also be used as the function g in the estimate of $K_r(f,t)$ in terms of the modulus of smoothness. The Jackson estimates handle the term $\|f-g\|_p$. The term $\|g^{(r)}\|_p$ is estimated in a way similar to the usual proofs of the Bernstein inverse theorem for approximation by trigonometric polynomials [22].

Let us point out some uses of the K-functional. Suppose that we want to approximate functions from $L_p(I)$ or $L_p(2\pi)$ by elements from some set A such as polynomials or splines. Then, it is enough by virtue of Theorem 2.1 to estimate the approximation for smooth functions $g \in W_{r,p}(I)$ or respectively $W_{r,p}(2\pi)$.

THEOREM 2.2. *Let $\varepsilon > 0$, $1 \le p \le \infty$. Suppose that for each $g \in W_{r,p}(I)$ there is an element $a_g \in A$ such that $\|g-a_g\|_p \le C_1\varepsilon^r\|g^{(r)}\|_p$, with $C_1 \ge 1$ depending only on r. Then, for each $f \in L_p(I)$, $1 \le p < \infty$, or $f \in C(I)$, $p=\infty$, there is an $a \in A$ such that*

$$\|f-a\|_p \le C_2\omega_{r,p}(f,\varepsilon),$$

with C_2 depending only on r. A similar result holds in the periodic case.

Proof. We consider only the case of $L_p(I)$. The other cases are similar. If $f \in L_p(I)$ and g is any function in $W_{r,p}(I)$ with a_g as in the hypotheses, then

$$\|f-a_g\|_p \le \|f-g\|_p + \|g-a_g\|_p \le C_1\{\|f-g\|_p + \varepsilon^r \|g^{(r)}\|_p\}.$$

By the definition of the K-functional, we can choose g appropriately so that

$$\|f-a_g\|_p \le 2C_1K_{r,p}(f,\varepsilon^r) \le C_2\omega_{r,p}(f,\varepsilon),$$

where we have used Theorem 2.1 in our last inequality.

If the approximation is given by a linear operator L, then Lf will provide the desired approximation.

THEOREM 2.3. *Let* $1\le p\le\infty$. *Suppose* L *is a bounded linear operator from* $L_p(I)$ $(C(I), p=\infty)$ *into itself. If* $\varepsilon>0$ *and for each* g *in* $W_{r,p}(I)$ *we have* $\|g-Lg\|_p \le C_1\varepsilon^r \|g^{(r)}\|_p$ *with* $C_1\ge 1$ *depending only on* r. *Then for each* $f\varepsilon L_p(I)$ $(C(I), p=\infty)$ *we have*

$$\|f-Lf\|_p \le C_2\omega_{r,p}(f,\varepsilon),$$

with C_2 *a constant depending only on* r *and* $\|L\|$. *A similar result holds in the periodic case.*

Proof. We again only consider $L_p(I)$, $1\le p<\infty$. The other cases are the same. If $f\varepsilon L_p(I)$ and g is any function in $W_{r,p}(I)$, we have

$$\begin{aligned}\|f-Lf\|_p &\le \|f-g\|_p + \|g-Lg\|_p + \|Lf-Lg\|_p\\ &\le C_1(1+\|L\|)\{\|f-g\|_p + \varepsilon^r\|g^{(r)}\|_p\}.\end{aligned}$$

Taking now an infimum over all g and using the definition of the K-functional as well as Theorem 2.1, we find

$$\|f-Lf\|_p \le CK_{r,p}(f,\varepsilon^r) \le C_2\omega_{r,p}(f,\varepsilon),$$

with C_2 depending only on r and $\|L\|$.

When the approximation is given by a linear operator and $p=\infty$, we can even give pointwise estimates.

THEOREM 2.4. *Suppose* L *is a bounded linear operator from* C(I) *into itself and* $\varepsilon(x)\ge 0$, $-1<x\le 1$. *If* $C_1\ge 1$ *is such that for each* $g\varepsilon W_{r,\infty}(I)$, *we have*

$$|g(x)-L(g,x)| \le C_1\|g^{(r)}\|_\infty(\varepsilon(x))^r, \quad -1\le x\le 1.$$

Then for each $f\varepsilon C(I)$, *we have*

$$|f(x)-L(f,x)| \leq C_2\omega_r(f,\varepsilon(x)), \quad -1\leq x\leq 1,$$

<u>with</u> C_2 <u>depending only on</u> r <u>and</u> $\|L\|$. <u>A similar result holds for</u> $C(2\pi)$.

<u>Proof</u>. Fix x and consider $f\varepsilon C(I)$ and g an arbitrary function in $W_{r,\infty}(I)$. Then,

$$\begin{aligned}|f(x)-L(f,x)| &\leq |f(x)-g(x)| + |g(x)-L(g,x)| + |L(f-g,x)| \\ &\leq (1+\|L\|)\|f-g\|_\infty + C_1\|g^{(r)}\|_\infty(\varepsilon(x))^r \\ &\leq C_1(1+\|L\|)\{\|f-g\|_\infty + (\varepsilon(x))^r\|g^{(r)}\|_\infty\}.\end{aligned}$$

Taking now an infimum over all such g and using the definition of the K- functional and Theorem 2.1, we find

$$|f(x)-L(f,x)| \leq CK_r(f,(\varepsilon(x))^r) \leq C_2\omega_r(f,\varepsilon(x)),$$

with C_2 depending only on r and $\|L\|$. Since x was arbitrary, we have proved the theorem.

These last three theorems will enable us in what follows to just prove our estimates for smooth functions. We would have to do this in any case but in most instances this is made simpler with added smoothness.

Let us give another easy application of Theorem 2.1 by showing how a function can be extended from one interval to a larger interval with no loss of smoothness as measure by ω_r. The method of extension was given by H. Whitney[33]. Let J= [0,1] and as usual I=[-1,1]. We will examine the extension from J to I.

If $f\varepsilon L_p(J)$ (C(J)), then define

$$(2.18)\quad Tf(x) = \begin{cases} f(x), & x\varepsilon J \\ \sum_0^r c_i f(-2^{-i}x), & x\varepsilon I-J, \end{cases}$$

where the numbers c_i satisfy

(2.19) $\sum_{i=0}^{r} c_i(-2^{-i})^j = 1, \quad j=0,1,\ldots,r.$

The function Tf is an extension of f to $L_p(I)$ $(C(I))$. It is clear that T is a bounded linear operator from $L_p(J)$ to $L_p(I)$ $(C(J)$ to $C(I))$. The condition (2.19) guarantees that polynomials of degree r are extended to the same polynomial.

If $g \in W_{r,p}(J)$, then $Tg \in W_{r,p}(I)$ since the only point in question is x=0, where we have continuity for each function $(Tg)^{(j)}$, $j=0,1,\ldots,r-1$ because of (2.19). Also, it is clear that $\|(Tg)^{(r)}\|_p(I) \leq C\|g^{(r)}\|_p(J)$ with C depending only on r. Hence, if f is in $L_p(J)$ $(C(J))$ and g is an arbitrary function in $W_{r,p}(J)$, then for $t>0$,

$$\begin{aligned}\|\Delta_t^r(Tf,x)\|_p(I) &\leq \|\Delta_t^r(T(f-g),x)\|_p(I) + \|\Delta_t^r(Tg,x)\|_p(I)\\ &\leq 2^r\|T(f-g)\|_p(I) + t^r\|(Tg)^{(r)}\|_p(I)\\ &\leq C\{\|f-g\|_p(J) + t^r\|g^{(r)}\|_p(J)\}.\end{aligned}$$

Taking now an infimum over all such g on the right side and a supremum over all $t\leq h$, we have from Theorem 2.1 that

$$\omega_{r,p}(Tf,h) \leq CK_{r,p}(f,h^r) \leq C\omega_{r,p}(f,h).$$

This shows that T preserves the smoothness of f.

THEOREM 2.5. *If* $[a,b] \subseteq [c,d]$, $r\geq 1$, *and* $1\leq p\leq\infty$, *then there is a linear operator* T *mapping* $L_p[a,b]$ *into* $L_p[c,d]$, $1\leq p<\infty$, *and* $C[a,b]$ *into* $C[c,d]$, $p=\infty$, *such that when* $f\in L_p[a,b]$ $(f\in C[a,b])$, *then* $Tf(x) = f(x)$, $x\in[a,b]$ *and*

(2.20) $\omega_{r,p}(Tf,h) \leq C\omega_{r,p}(f,h), \quad h>0,$

with C *a constant depending only on* r *and the two intervals.*

Proof. We have seen this theorem for extending from $[0,1]$ to $[-1,1]$. The same idea gives an extension from $[-1,0]$ to $[-1,1]$. These two results can then be combined with usual linear change of scale to give the general result.

We have seen a few ways that the K- functional can be used for approximation related questions. The first general description of the role of the K-functional in approximation was given by J. Peetre [27] (see also [28]). P. L. Butzer - H. Berens [7] and H. Berens [2] develop the uses of the K-functional in problems relating to the degree of approximation , especially for semi-groups of operators. A general treatment for direct and inverse theorems of approximation is given by P. L. Butzer - K. Scherer [8]. The K-functional seems to have had some of its earliest and most powerful uses in spline approximation as for example already in R. Varga [35] and K. Scherer [31]. We will see many more uses of the K-functional approach in the following sections.

3 Local Approximation by Polynomials of Fixed Degree

We want to estimate the error in approximating f on an interval $[a,b]$ I by polynomials of degree $\leq r-1$ in terms of the r-th order modulus of smoothness. Thus, in distinction to the Jackson theorems, here the degree is fixed. The importance of obtaining such estimates is twofold. First, spline approximation can be viewed as a smooth joining together of local polynomial approximations. This means our estimates will be useful in the next section. But beyond the spline approximation, local polynomial approximation will give another view of the connections between degree of approximation and smoothness.

THEOREM 3.1. *If* $f \in L_p[a,b]$ *and* $\delta = \frac{1}{8} r^{-2}|b-a|$, *then there is a polynomial* Q *of degree* r-1, *such that*

$$(3.1) \quad \|f-Q\|_p[a,b] \leq C\{\int_0^r \|\Delta^r_{\delta u}(f,x)\|_p(J_{r\delta u})du$$

$$+ \max_{1\leq k\leq r} \|\Delta^r_{\delta k}(f,x)\|_p(J_{r\delta k})\} \leq C\omega_{r,p}(f,|b-a|),$$

with $J_t = [a,b-t]$, $J = [a,b]$, *and* C *a constant depending only*

on r.

Proof. It is enough to prove (3.1) for $[a,b] = I$. The general case follows by the usual change of variables. So $\delta = \frac{1}{4} r^{-2}$. Suppose that $f \in L_p(I)$ and let g be as in Lemma 2.2 for $t=\delta$. Consider the polynomial $Q(x) = g(0)+g'(0)x+\cdots+(g^{(r-1)}(0)/(r-1)!)\cdot x^{r-1}$, which is the first r terms of the Taylor expansion of g about 0. For each $1 \le i \le r$ and $-1 \le x \le 1$, we have

$$|g^{(i-1)}(x)-Q^{(i-1)}(x)| = |\int_0^x (g^{(i)}(u)-Q^{(i)}(u))du|$$
$$\le \|g^{(i)}-Q^{(i)}\|_p,$$

because of Hölder's inequality. Hence, $\|g^{(i-1)}-Q^{(i-1)}\|_p \le \|g^{(i)}-Q^{(i)}\|_p$, $i=1,2,\ldots,r$. This gives the estimate

(3.2) $\|g-Q\|_p \le \|g^{(r)}\|_p \le C\delta^r \|g^{(r)}\|_p$.

Therefore, if we estimate $\|f-Q\|_p \le \|f-g\|_p + \|g-Q\|_p$ and use (2.16) on the first term and (2.17) along with (3.2) on the second term, we obtain the estimate (3.1) as desired.

The middle estimate given in (3.1) is more sensitive than $\omega_{r,p}(f,|b-a|)$ in the sense that we can add up such estimates and still retain an inequality of the same form. The exceptional case is $p=\infty$ where the modulus of smoothness suffices.

In the case $p=\infty$, Theorem 3.1 was given by H. Whitney [34]. Whitney's proof is more complicated but includes a sensitive estimate of the constants. The proof we have given is made simple by the use of the K-functional.

The estimate (3.1) is of course an estimate for the degree of approximation in terms of smoothness. Actually, these estimates can be easily reversed. Consider the simplest case when $p=\infty$. Let $E_{r-1}(f,J)$ denote the error in approximating $f \in C(J)$ in $\|\cdot\|_\infty(J)$ by polynomials of degree r-1. From Theorem 3.1 it follows that if $t>0$, then

(3.3) $\sup\{ E_{r-1}(f,J) : J \subset I, |J| \le rt\} \le C\omega_r(f,t)$

The reverse inequality also holds, since if $t>0$ and $[x,x+rt]\subset I$, with $0\le h\le t$, $J=[x,x+rh]$ and Q any polynomial of degree $r-1$, then we have

$$|\Delta_h^r(f,x)| = |\Delta_h^r(f-Q,x)| \le 2^r\|f-Q\|_\infty(J).$$

Taking an infimum over the Q's on the right hand side and a supremum over $x\in I$, $h\le t$ on the left hand side gives the converse to (3.3).

If $J=[a,b]$ and P_r is any bounded projection from $L_p(J)$ into the space of polynomials of degree $\le r-1$, then the following is a simple consequence of Theorem 3.1.

COROLLARY 3.1. <u>There is a constant</u> C <u>depending only on</u> r, <u>so that if</u> $\delta=\frac{1}{8}r^{-2}|b-a|$ <u>and</u> $f\in L_p(J)$, <u>then for any projection</u> P_r,

$$\|f-P_r(f)\|_p(J) \le C(1+\|P_r\|)\Big\{\int_0^r \|\Delta_{\delta u}^r(f,x)\|_p(J_{r\delta u})du$$

$$+\max_{1\le k\le r}\|\Delta_{\delta k}^r(f,x)\|_p(J_{r\delta k})\Big\} \le C\omega_{r,p}(f,|b-a|),$$

<u>with</u> $J_t=[a,b-t]$, $J=[a,b]$, <u>and</u> C <u>depending only on</u> r and $\|P_r\|$.

<u>Proof</u>. If Q is any polynomial of degree $r-1$, we have

$$\begin{aligned}\|f-P_r(f)\|_p(J) &\le \|f-Q\|_p(J) + \|P_r(f-Q)\|_p(J)\\ &\le (1+\|P_r\|)\|f-Q\|_p(J).\end{aligned}$$

If we now take Q to satisfy Theorem 3.1, we obtain the desired estimate.

4 <u>Approximation by Splines</u>

Let $\Pi=\{x_i\}_0^n$ be a set of knots with $-1=x_0<x_1<\cdots<x_n=1$. If $r\ge1$, then denote by $\mathcal{S}_r(\Pi)$ the space of splines of order r and knots Π. Thus $S\in\mathcal{S}_r(\Pi)$ if and only if $S^{(r-2)}$ is continuous and on each interval $(x_i,x_{i+1}]$, S is a polynomial of degree $\le r-1$. When $r=1$, we assume left continuity at each knot and right continuity at -1.

We will estimate the degree of approximation by splines from $\mathcal{S}_r(\Pi)$. Our estimates automatically include estimates for splines with multiple knots since the latter splines are less smooth. In spline approximation, it is customary to measure the degree of approximation in terms of the mesh ratio $|\Pi| = \max(x_{i+1}-x_i)$, while holding the order r fixed.

If on the other hand the knots are held fixed and the degree tends to infinity then, this is akin to polynomial approximation. When both r and n tend to infinity, then one can to a certain extent superimpose the results from the other two cases, but care must be taken in the estimation of the constants. We will only consider the case where r is fixed.

An important role in the construction of spline approximants is played by the B-splines. These splines have minimal support and are a basis for $\mathcal{S}_r(\Pi)$. Define $x_i = -1-i(x_1+1)$, $i<0$, and $x_i=1+(1-x_{n-1})$, $i>n$. This extends the knots with the same sort of spacing. Let $g_r(s;x) = (s-x)_+^{r-1}$ and for $-r<i<n$, define

$$(4.1)\quad N_{i,r}(x) = (x_{i+r}-x_i)g_r(x_i,\ldots,x_{i+r};x),$$

with the notation meaning that we are taking an r-th divided difference with respect to the variable s while holding x fixed. The spline $N_{i,r}$ is non-negative and vanishes outside of (x_i,x_{i+r}). The B-splines $N_{i,r}$, $-r<i<n$, form a basis for $\mathcal{S}_r(\Pi)$ and with our normalization, we have $\sum_{-r<i<n} N_{i,r}(x) = 1$, $x\in I$.

To construct spline approximants, we can follow our general plan. In other words, we want first to construct approximants to smooth function $g\in W_{r,p}(I)$. For such a function g, we will show that there are splines S_j, $j=1,\ldots,r$, such that $S_j\in\mathcal{S}_j(\Pi)$,

$$\|g^{(r-j)}-S_j\|_p \le (r|\Pi|)^j\|g^{(r)}\|_p.$$

Of course, we need only the case j=r. but the intermediate re-

sults are used since the proof goes by induction.

Let us sketch the construction which follows the ideas of B. Popov and Bl. Sendov [29]. Subtracting a polynomial of degree r-1 if necessary, we can assume that $g^{(i)}(-1) = 0$, $i=0,\dots,r-1$. Let S_1 be the step function in $\mathcal{S}_1(\Pi)$ with $S_1(x) = g^{(r-1)}(x_i)$, $x_i < x \le x_{i+1}$. Then, by Hölder's inequality,

$$(4.2)\quad |g^{(r-1)}-S_1(x)|^p \le \left\{ \int_{x_i}^{x_{i+1}} |g^{(r)}(t)|\,dt \right\}^p \le |x_{i+1}-x_i|^{p/q} \int_{x_i}^{x_{i+1}} |g^{(r)}(t)|^p dt,$$

when $x_i \le x \le x_{i+1}$. Here q is the conjugate index to p, $1/p+1/q=1$. Integrating this last inequality over $[x_i, x_{i+1}]$, replacing $|x_{i+1}-x_i|$ by $|\Pi|$ and then summing over i and taking a p-th root gives $\|g^{(r-1)}-S_1\|_p \le \|g^{(r)}\|_p\,|\Pi|$, which is the case $j=1$.

Now to the induction step. Suppose that $S_j \varepsilon \mathcal{S}_j(\Pi)$ satisfies $\|g^{(r-j)}-S_j\|_p \le (r|\Pi|)^j\,\|g^{(r)}\|_p$. Let $\lambda=[n/r]$ and $y_i=x_{ri}$, $i=0,\dots,\lambda$. Define for $0\le i<\lambda$,

$$a_{i,j} = \int_{y_i}^{y_{i+1}} \{g^{(r-j)}(t)-S_j(t)\}dt, \qquad c_{i,j}^{-1} = \int_{y_i}^{y_{i+1}} N_{ri,j}(t)dt.$$

Note that $N_{ri,j}$ is supported on $(x_{ri}, x_{ri+j}) \subseteq (y_i, y_{i+1})$. The spline S_{j+1} defined by

$$S_{j+1}(x) = \int_{-1}^{x} \{S_j(t) + \sum_{i=o}^{\lambda-1} a_{i,j}\, c_{i,j}\, N_{ri,j}(t)\}\, dt$$

is in $\mathcal{S}_{j+1}(\Pi)$ and satisfies $\|g^{(r-j-1)}-S_{j+1}\|_p \le (r|\Pi|)^{j+1}\cdot \|g^{(r)}\|_p$. This is shown in the same way we have estimated (4.2) above, except that now we use the fact that $S_j(y_i) = g^{(r-j-1)}(y_i)$, $0\le i\le\lambda$, and $|y_{i+1}-y_i| \le r|\Pi|$, as well.

The above construction shows that for each $g\varepsilon W_{r,p}(I)$, there is a spline $S\varepsilon\,\mathcal{S}_r(\Pi)$ such that

(4.3) $\|g-S\|_p \le (r|\Pi|)^r \|g^{(r)}\|_p$.

This is the desired result for smooth functions.

THEOREM 4.1. For each $f \in L_p(I)$ there is an $S \in \mathcal{S}_r(\Pi)$ such that

(4.4) $\|f-S\|_p \le C\omega_{r,p}(f,|\Pi|)$,

with C a constant depending only on r.

Proof. This follows from Theorem 2.2 and the estimate (4.3).

There is another approach to obtaining spline approximants given by C. deBoor - G. Fix [3], which we also want to look at. Their technique yields linear projections from $L_p(I)$ to $\mathcal{S}_r(\Pi)$ which give the estimate (4.4).

We have already mentioned that the B-splines form a basis for $\mathcal{S}_r(\Pi)$. Thus any spline $S \in \mathcal{S}_r(\Pi)$ can be represented as $\sum a_i(S)N_{i,r}$, with the a_i's linear functionals on $\mathcal{S}_r(\Pi)$ and $a_i(N_{j,r}) = \delta_{i,j}$ (Kronecker delta notation). There are explicit formulae for the a_i's given in [4]. For example, if $\tau_j \in (x_j, x_{j+r})$, $\Psi_j(x) = (x_{j+1}-x)\cdots(x_{j+r-1}-x)$ and

$$(r-1)!\eta_{j,k} = (-1)^{r-k-1}\,\Psi_j^{(r-k-1)}(\tau_j),$$

then the functional a_j, $-r<j<n$, can be represented by

(4.5) $a_j(S) = \sum_{k<r} \eta_{j,k} S^{(k)}(\tau_j)$.

It is clear from (4.5) that the functionals a_j have support on (x_j, x_{j+r}). We want to bring this out more clearly in a manner in which we work with S and not its derivatives. We are still at liberty to choose the τ_j's. For $-r<j<n$, let I_j be a largest interval of the form (x_i, x_{i+1}) contained in $I \cap (x_j, x_{j+r})$. In case of a tie we choose I_j as the left most such interval. Our selection gives that $I_j \subset I$ and $r|I_j| \ge |x_{j+r}-x_j|$, $-r<j<n$, since the knots have a uniform spacing outside of I.

For each j, let τ_j be the midpoint of I_j.

LEMMA 4.1. Let $1 \le p \le \infty$. There is a constant C that depends only on r such that for each $S \in \mathcal{S}_r(\Pi)$, we have

$$(4.6) \quad |a_j(S)| \le C|x_{j+r}-x_j|^{-1/p} \|S\|_p(I_j), \quad -r<j<n.$$

Proof. First observe that there is a constant C depending only on r such that for any polynomial Q of degree $\le$ r-1, we have $\|Q\|_\infty \le C\|Q\|_1 \le C\|Q\|_p$. This is because on a finite dimensional space any two norms are equivalent. Changing this inequality to (a,b) gives

$$(4.7) \quad \|Q\|_\infty(a,b) \le C|b-a|^{-1/p} \|Q\|_p(a,b).$$

Now consider the representation (4.5) for a_j. From the formula for the $\eta_{j,k}$'s, it follows that

$$(4.8) \quad |\eta_{j,k}| \le (x_{j+r}-x_j)^k, \quad -r<j<n,\ 0\le k<r.$$

Hence, if we use Markov's inequality with (4.8) and (4.5), we find

$$|a_j(S)| \le \sum_{k=o}^{r-1} (x_{j+r}-x_j)^k \ |S^{(k)}(\tau_j)|$$

$$\le C\|S\|_\infty(I_j) \sum_{k=o}^{r-1} |x_{j+r}-x_j|^k |I_j|^{-k} \le C\|S\|_\infty(I_j),$$

with C depending only on r. Recall that $|x_{j+r}-x_j| \le r|I_j|$. Using (4.7) to write this last inequality in terms of the L_p norm, we arrive at the desired result (4.6).

The functionals a_j are only defined on $\mathcal{S}_r(\Pi)$ but we can use the Hahn-Banach theorem to extend these functionals to all of $L_p(I)$ ($C(I)$, $p=\infty$) with the preservation of the estimate (4.6). Because a_j is supported on I_j, it has a norm preserving extension $\bar{a}_j$ to $L_p(I_j)$. Let χ_j denote the characteristic function of I_j and define $\alpha_j(f) = \bar{a}_j(f\cdot\chi_j)$ for each $f \in L_p(I)$. Then for any spline S, we still have that $\alpha_j(S) = \bar{a}_j(S\cdot\chi_j) = a_j(S)$.

For simplicity of notation, we do not indicate the dependence of the extended functional α_j on p.

The functionals α_j will now satisfy for $f \in L_p(I)$ ($f \in C(I)$, $p=\infty$),

(4.9) $|\alpha_j(f)| \leq C|x_{j+r}-x_j|^{-1/p} \|f\|_p(I_j)$, $\quad -r<j<n$.

Define now the operators

(4.10) $L_{\Pi,p}(f\ x) = \sum\limits_{-r<i<n} \alpha_i(f)N_{i,r}(x)$, $\quad f \in L_p(I)(f \in C(I), p=\infty)$.

These are projections from $L_p(I)$ onto $\mathcal{S}_r(\Pi)$. These operators, since they are projections, are as effective as any spline approximation provided there norms are controllable. So our next task is to check the norm of $L_{\Pi,p}$.

If $0 \leq j<n$, then only the indices i with $j-r<i \leq j$ contribute to $L_{\Pi,p}(f,x)$ when $x_j \leq x \leq x_{j+1}$. Also, the $N_{i,r}$'s are non-negative and sum to one. Therefore, for $1 \leq p<\infty$,

$$
\begin{aligned}
(4.11) \quad \int_{x_j}^{x_{j+1}} |L_{\Pi,p}(f,x)|^p dx &\leq \max_{j-r<i\leq j} |\alpha_i(f)|^p \int_{x_j}^{x_{j+1}} dx \\
&\leq C \max_{j-r<i\leq j} \left\{|x_{i+r}-x_i|^{-1} \int_{I_i} |f(x)|dx \right\} |x_{j+1}-x_j| \\
&\leq C \max_{j-r<i\leq j} \int_{I_i} |f(x)|^p dx,
\end{aligned}
$$

with C depending only on r. Here, we have used our estimate for α_i given in (4.9) and the fact that $|x_{j+1}-x_j| \leq |x_{i+r}-x_i|$, $j-r<i \leq j$.

Now, for each j, the max in (4.11) is taken only over r terms and for any given interval (x_k, x_{k+1}), this interval can appear at most r times in the sequence of intervals (I_j). Therefore, we can sum the inequalities (4.11) over $0 \leq j \leq n$ and find

(4.12) $\|L_{\Pi,p}(f)\|_p \leq C\,\|f\|_p$, $f \in L_p(I)$ $(f \in C(I),\ p=\infty)$.

We have actually only proved the case $1 \leq p < \infty$ in (4.12) but the case $p=\infty$ is the same sort of argument. From (4.12), we have

(4.13) $\|L_{\Pi,p}\| \leq C$,

with C a constant depending only on r.

THEOREM 4.2. <u>Let</u> $1 \leq p \leq \infty$, <u>and define</u> $L_{\Pi,p}$ <u>as in</u> (4.10). <u>Then, there is a constant</u> C <u>depending only on</u> r <u>such that</u>

(4.14) $\|f-L_{\Pi,p}(f)\|_p \leq C\omega_{r,p}(f,|\Pi|)$, $f \in L_p(I)$ $(f \in C(I),\ p=\infty)$.

<u>Proof</u>. If S is any spline in $\mathcal{S}_r(\Pi)$, then since $L_{\Pi,p}$ is a projection, we have

$$\|f-L_{\Pi,p}(f)\|_p \leq \|f-S\|_p + \|L_{\Pi,p}(f-S)\|_p \leq (1+\|L_{\Pi,p}\|)\,\|f-S\|_p.$$

Now, taking S as in Theorem 4.1 and using (4.13), we have the desired result (4.14).

Our proof of Theorem 4.2 relied on Theorem 4.1 and hence in turn on Theorem 2.1. However, it is possible to prove Theorem 4.2 directly by using the estimates of Theorem 3.1. In this way, we can get another proof of Theorem 2.1. Namely, take n equally spaced knots for the knot set Π_n. If $(n+1)^{-1} \leq t \leq n^{-1}$, let $g = L_{\Pi_n,p}(f)$ for r+1 in place of r. Then g will give the crucial right hand estimate in (2.7) (see [15]).

In the case $p=\infty$, there is an additional benefit to arguing directly since then we can get local estimates for the approximation as shown in the next theorem.

THEOREM 4.3. <u>Let</u> $L_\Pi = L_{\Pi,\infty}$ <u>as in</u> (4.10). <u>If</u> $f \in C(I)$ <u>and</u> $x_j \leq x \leq x_{j+1}$, $0 \leq j < n$, then

(4.15) $|f(x)-L_\Pi(f,x)| \leq C\omega_r(f,|x_{j+r}-x_{j-r}|)$,

<u>with</u> C <u>a constant depending only on</u> r.

<u>Proof</u>. Let Q be a polynomial which satisfies (3.1) for $p=\infty$

and $[a,b] = [x_{j-r}, x_{j+r}]$. Then,

$$(4.16)\quad \|f-Q\|_\infty[x_{j-r},x_{j+r}] \le C\omega_r(f,|x_{j+r}-x_{j-r}|).$$

Now, $N_{i,r}(x) = 0$ if $i \notin [j-r+1,j]$, and the $N_{i,r}$'s sum to one. Hence, if $x_j \le x \le x_{j+1}$,

$$(4.17)\quad |L_\Pi(f-Q,x)| \le \max_{j-r<i\le j} |\alpha_i(f-Q)| \le C \max_{j-r<i\le j} \|f-Q\|_\infty(I_i)$$
$$\le C\,\|f-Q\|_\infty[x_{j-r},x_{j+r}],$$

because of (4.9). Since L_Π preserves polynomials of degree r-1, we have from (4.16) and (4.17)

$$|L_\Pi(f,x)-f(x)| \le |L_\Pi(f-Q,x)| + |f(x)-Q(x)|$$
$$\le C\,\|f-Q\|_\infty[x_{j-r},x_{j+r}] \le C\omega_r(f,|x_{j+r}-x_{j-r}|),$$

with C depending only on r. Here, we used the fact that L_Π is bounded and (4.16). This is the desired estimate (4.15).

Let us point out one special case. Let $y_i = 1-n^{-2}(n-i)^2$, $i=0,1,\ldots,n$ and $y_{-i}=y_i$, $i=1,2,\ldots,n$. Define the knot set $\Pi=\{x_i\}_0^{2n}$, with $x_i = y_{i-n}$, $i=0,1,\ldots,2n$. These knots are thicker near the end points. If we let $\Delta_n(x) = \max\{n^{-1}(1-x^2)^{\frac{1}{2}}, n^{-2}\}$, then from Theorem 4.3 it follows that

$$(4.18)\quad |f(x)-L_\Pi(f,x)| \le C\omega_r(f,\Delta_n(x)),\quad -1\le x\le 1.$$

This estimate is a spline analogue of the Timan type estimates for algebraic polynomial approximation (see Section 7). For splines, such pointwise improvements can also be made in the interior of the interval by allowing the knots to thicken near a point or a finite number of points. Such interior improvements are not possible in polynomial approximation [11,p.190].

There is a much larger selection of coefficient functionals α_j which when used in the definition of $L_{\Pi,p}$ will yield the Jackson order estimates of Theorems 4.1 and 4.2. A careful analysis of the proof given for $p=\infty$ in Theorem 4.3 shows that

we do not need that $L_{\Pi,p}$ is a projection but it suffices that polynomials of degree $\leq$ r-1 are preserved and estimates like (4.6) and (4.9) hold. As we have mentioned already, Theorem 4.2 can be proved directly for general p and here again the direct proof requires only that $L_{\Pi,p}$ preserve polynomials of degree $\leq$ r-1 and the inequalities for α_i. This leads to a greater variety of possible choices for the coefficient functionals. These ideas are examined in detail in T. Lyche - L. Schumaker [25].

5 Approximation by Trigonometric Polynomials

There are several ways of deriving direct estimates for approximation by trigonometric polynomials. Let us consider only the most common method which is to use some sort of convolution type operator. Suppose that (K_n) is a sequence of non-negative trigonometric polynomials of degree n and normalized to have integral over $[-\pi,\pi]$ equal to one. The operators L_n defined by

$$(5.1)\quad L_n(f,\theta) = \int_{-\pi}^{\pi}\{(-1)^{r+1}\,\Delta_t^r(f,\theta) + f(\theta)\}K_n(t)dt,\ f\varepsilon L_p(2\pi)$$

will yield the Jackson estimates provided that the moments of the kernels satisfy certain inequalities.

We begin by checking smooth functions in $W_{r,p}(2\pi)$. When $g\varepsilon W_{r,p}(2\pi)$ and $M = \|g^{(r)}\|_p$, then

$$(5.2)\quad \|g-L_n(g)\|_p \leq \int_{-\pi}^{\pi}\|\Delta_t^r(g,\theta)\|_p K_n(t)dt \leq M\int_{-\pi}^{\pi}|t|^r K_n(t)dt\,.$$

Therefore, to have the correct order of estimates for functions in $W_{r,p}(2\pi)$, we need that the r-th moments of the K_n's be of order $O(n^{-r})$.

The best known examples of kernels satisfying this moment condition are the Jackson kernels $J_n(t) = c_{n,r}(\sin(mt/2))^{2r}\cdot(\sin(t/2))^{-2r}$, with $m = [n/r]$ and $c_{n,r}$ a normalizing constant. The kernels J_n satisfy (see [22, p. 57])

$$(5.3)\qquad \int_{-\pi}^{\pi} |t|^r J_n(t)dt \leq Cn^{-r},$$

with C a constant depending only on r.

THEOREM 5.1. *Let $1 \leq p \leq \infty$ and (K_n) be a sequence of non-negative trigonometric polynomials of degree $\leq n$ which satisfy*

$$(5.4)\qquad \int_{-\pi}^{\pi} |t|^r K_n(t)dt \leq Cn^{-r}, \quad n = 1,2,\ldots$$

with C a constant depending only on r. Define the operators L_n as in (5.1). Then for each $n \geq 1$ and $f \epsilon L_p(2\pi)$, $L_n(f)$ is a trigonometric polynomial of degree $\leq n$ for which

$$(5.5)\qquad \| f-L_n(f)\|_p \leq C\omega_{r,p}(f,n^{-1}),$$

with C a constant depending only on r.

Proof. Because of (5.4) and (5.2), for any $g \epsilon W_{r,p}(2\pi)$, we have $\| g-L_n(g)\|_p \leq Cn^{-r}\| g^{(r)}\|_p$. Thus the theorem follows from Theorem 2.3, since each L_n clearly has norm $\leq 2^r$.

6 Approximation by algebraic polynomials

The usual technique for proving Jackson type estimates for approximation by algebraic polynomials when $p=\infty$ is to revert back to the trigonometric case by using the substitution $x = \cos\theta$. This is particularily useful for pointwise estimates as described in the next section. However, in the L_p case, $1\leq p<\infty$, this method is not satisfactory since when $f\epsilon L_p(I)$, the function $g(\theta) = f(\cos\theta)$ need not be in $L_p(2\pi)$ (e.g. $f(x)=(1-x^2)^{-\alpha}$, $\alpha= 3/4p$). Instead, we will work to construct algebraic analogues of the operators in Section 5.

Let us first note that there are algebraic polynomials μ_n of degree $\leq n$ which are non-negative and satisfy

$$(6.1)\qquad \int_{-1/r}^{1/r} \mu_n(t)dt = 1,$$

$$(6.2)\qquad \int_{-3}^{3} |t|^r \mu_n(t)dt \le Cn^{-r}$$

with C a constant depending only on r. Examples of how to construct such polynomials for r=2 were given in [10] and [11]. The same approach carries over to general r. A sketch of the construction goes as follows.

Let P_{2m} be the Legendre polynomial of degreee 2m and $x_{i,m}$, i=1,...,m its positive zeros written in increasing order. The polynomials

$$\lambda_n(t) = c_n \left(\frac{P_{2m}(t)}{(x^2-x_{1,m}^2)\cdots(x^2-x_{r,m}^2)} \right)^2$$

with m = [n/4] have degree $\le$ n and with a proper choice of c_n,

$$\int_{-1}^{1} \lambda_n(t)dt = 1, \qquad \int_{-1}^{1} |t|^r \lambda_n(t)dt \le Cn^{-r}.$$

The estimate of the r-th moment of λ_n can be done by the Gauss quadrature formula when r is even and by using the Cauchy-Schwartz inequality and then the Gauss quadrature formula when n is odd. Now switching to [-3,3] and adjusting the constant so that (6.1) holds will give polynomials that satisfy (6.1) and (6.2). Note that the integral of λ_n on [-1,1] - [-1/r,1/r] is $O(n^{-r})$ because of the moment inequality.

Now to the construction of the operators. Because the functions are not periodic, we need to work with some sort of extension. We can use the extension introduced at the end of Section 2. Take [a,b] = [-1,1] and [c,d] = [-4,4] and let T be the linear operator guaranteed by Theorem 2.5. For simplicity of notation, we write $\underline{f}$ in place of Tf.

For each $1\le k\le r$ and $-1\le x\le 1$,

$$(6.3)\qquad \int_{-2}^{2} \underline{f}(t)\mu_n(k^{-1}(t-x))k^{-1}dt = \int_{(-2-x)/k}^{(2-x)/k} \underline{f}(x+kt)\mu_n(t)dt$$

is an algebraic polynomial of degree $\le n$. Although the limits of integration of the term on the right hand side are variable, we shall soon see that only the integral over $[-\delta,\delta]$, with $\delta = 1/r$ is significant.

The analogue of the trigonometric operators (5.1) are

$$(6.4)\quad \Phi_n(f,x)= \int_{-2}^{2} \underline{f}(t)\phi_n(t-x),\quad \phi_n(u)= -\sum_{1}^{r} (-1)^k\binom{r}{k}k^{-1}\mu_n(k^{-1}u).$$

Using (6.3), $\Phi_n(f)$ can be written in terms of the r-th difference of $\underline{f}$. For each $1\le k\le r$, $-1\le x\le 1$, we have $(-2-x)/k \le -\delta \le \delta \le (2-x)/k$. Thus the right hand integral in (6.3) can be written as an integral over $[-\delta,\delta]$ and a remainder $R_k(x)$. Hence,

$$(6.5)\quad \Phi_n(f,x) = \int_{-\delta}^{\delta} \{(-1)^{r+1}\,\Delta_t^r(f,x) + f(x)\}\,\mu_n(t)dt$$
$$+ \sum_{1}^{r} (-1)^{k+1}\binom{r}{k}R_k(x) \;=\; A_n(f,x) + B_n(f,x).$$

The operator A_n is now of a form akin to (5.1). We want now to estimate B_n. Let $J_{x,k} = [(-2-x)/k,(2-x)/k] - [-\delta,\delta]$, and $J = \{t\colon \delta\le|t|\le 3\}$. For each $1\le k\le r$ and $-1\le x\le 1$, $J_{x,k}\subset J$ and so

$$\|R_k\|_p = \|\int_{J_{x,k}} \underline{f}(x+kt)\mu_n(t)dt\|_p \le \|\underline{f}\|_p[-4,4] \int_J \mu_n(t)dt$$
$$\le \|f\|_p \int_{-3}^{3} (r|t|)^r\mu_n(t)dt \le Cn^{-r}\,\|f\|_p,$$

with C a constant depending only on r. In the second to last inequality, we used the fact that $r|t| \ge 1$ if $|t| \ge\delta$, and the fact that $\|\underline{f}\|_p[-4,4] \le C\|f\|_p$, since T is a bounded operator. In the last inequality, we used our moment estimate (6.2). This shows that

$$(6.6)\quad \|B_n(f)\|_p \le C\,\|f\|_p\, n^{-r},\qquad f\in L_p(I).$$

We want now to check A_n. If $g \varepsilon W_{r,p}(I)$, then

$$(6.7)\quad \|g-A_n(g)\|_p \le \int_{-\delta}^{\delta} \|\Delta_t^r(\tilde g,x)\|_p \mu_n(t)dt$$

$$\le \|\tilde g^{(r)}\|_p[-2,2] \int_{-3}^{3} |t|^r \mu_n(t)dt \le C\,\|g^{(r)}\|_p n^{-r}.$$

Here we have used the fact that $\tilde g = g$ on I, $\|\tilde g^{(r)}\|_p[-2,2] \le \|\tilde g^{(r)}\|_p[-4,4] \le C\,\|g^{(r)}\|_p$, since T preserves smoothness.

LEMMA 6.1. *Let $1 \le p \le \infty$. If $n \ge 1$, define Φ_n by* (6.4). *For each $f \varepsilon L_p(I)$, $\Phi_n(f)$ is an algebraic polynomial of degree $\le n$ and*

$$(6.8)\quad \|f-\Phi_n(f)\|_p \le C\{\omega_{r,p}(f,n^{-1}) + \|f\|_p n^{-r}\},$$

with C a constant depending only on r.

Proof. Because of (6.7) and Theorem 2.3, for each $f \varepsilon L_p(I)$, we have $\|f-A_n(f)\|_p \le C\omega_{r,p}(f,n^{-1})$. We have already seen that $\|B_n(f)\|_p \le C\,\|f\|_p n^{-r}$ by (6.6). Since $\Phi_n = A_n + B_n$, these two facts prove the lemma.

We can remove the unpleasant term $\|f\|_p n^{-r}$ by a minor adjustment. Given f, let F be the primitive of f with $F(0)=0$ and let Q be the polynomial of degree r which interpolates F at the points $-1+2ir^{-1}$, $i=0,1,\dots,r$. Define the projection P_r from $L_p(I)$ onto polynomials of degree $\le r-1$ by $P_r(f) = Q'$. Consider now the operator

$$(6.9)\quad L_n(f) = \Phi_n(f-P_r(f)) + P_r(f).$$

For $n \ge r$, $L_n(f)$ is still a polynomial of degree $\le n$, but now L_n preserves polynomials of degree $r-1$.

THEOREM 6.1. *Let $1 \le p \le \infty$. Define L_n as in* (6.9). *Then, there is a constant C, depending only on r, such that for $n \ge r$, $f \varepsilon L_p(I)$, $L_n(f)$ is a polynomial of degree $\le n$ with*

$$(6.10)\quad \|f-L_n(f)\|_p \le C\omega_{r,p}(f,n^{-1}).$$

Proof. We have from the definition (6.9) and from (6.8) that

$$(6.11)\quad \|f-L_n(f)\|_p \le \|\Phi_n(f-P_r(f))-(f-P_r(f))\|_p$$
$$\le C\omega_{r,p}(f,n^{-1}) + Cn^{-r}\,\|f-P_r(f)\|_p .$$

The last term can be estimated by using Corollary 3.1 to find

$$n^{-r}\,\|f-P_r(f)\|_p \le Cn^{-r}\omega_{r,p}(f,2) \le C\omega_{r,p}(f,n^{-1}),$$

where we have used the usual properties of the modulus of smoothness. When this last estimate is replaced in (6.11), we obtain (6.10), as desired

In the case that $r=1$, J. Bak - D. Newman [1] have shown the existence of polynomials Q_n which when replaced for $L_n(f)$ will give (6.10). This is done in the more general context of Müntz-Jackson theorems. Suprisingly, I could find no reference to the general case of r.

7 Pointwise approximation by algebraic polynomials

It was first observed by A.F. Timan [32] that in the case $p=\infty$, the approximation of f by algebraic polynomials can be improved near the end points of the interval. These improvements are necessary if we want a direct-inverse theorem characterization of classes of functions, e.g. Lip α, $0<\alpha<1$, in terms of approximation by algebraic polynomials. J. Brudnyi [5] has extended Timan's results to arbitrary ω_r by showing that for each $f\in C(I)$, $r\ge 1$, $n\ge r$, there is an algebraic polynomial P of degree $\le n$ such that

$$(7.1)\quad |f(x)-P(x)| \le C\omega_r(f,\Delta_n(x)), \quad -1\le x\le 1,$$

with C a constant depending only on r and $\Delta_n(x) = \max\{n^{-1}\cdot(1-x^2)^{\frac{1}{2}}, n^{-2}\}$. The case $r=1$ of (7.1) is the original result of Timan, while $r=2$ was proved independently by G. Freud [18] and V. Dzadyk [16]. Actually, any of the trigonometric operators introduced in Section 5 lead to algebraic methods that give

(7.1).

If we want to establish (7.1) using a sequence (Λ_n) of algebraic polynomial operators, then according to Theorem 2.4, we only need that for each $g\epsilon W_{r,p}(I)$,

(7.2) $\quad |g(x)-\Lambda_n(g,x)| \leq C(\Delta_n(x))^r \|g^{(r)}\|_\infty, \quad -1\leq x\leq 1,$

with C a constant depending only on r.

Let (K_n) be a sequence of trigonometric polynomials with K_n of degree n and integral one over $[-\pi,\pi]$, $K_n\geq 0$ and

(7.3) $\quad \int_{-\pi}^{\pi} |t|^j K_n(t)dt \leq Cn^{-j}, \quad j=1,2,\ldots,2r.$

For example the K_n's that satisfy (5.4) with 2r in place of r will satisfy (7.3). Define L_n as

(7.4) $\quad L_n(h,\theta) = \int_{-\pi}^{\pi} \{-\Delta_t^{2r}(h,\theta) + h(\theta)\}K_n(t)dt, \quad h\epsilon C(2\pi).$

Denote by $P_r(f)$ the polynomial of degree r-1 which interpolates f at the points $-1+2ir^{-1}$, $i=1,\ldots,r$. Then for $f\epsilon C(I)$, define

(7.5) $\quad \Lambda_n(f,x) = L_n(f(\cos\theta)-P_r(f,\cos\theta), \cos^{-1}x) + P_r(f,x).$

The use of P_r is to guarantee that $\Lambda_n(Q) = Q$ whenever Q is a polynomial of degree $\leq$ r-1. These operators will satisfy (7.2) as we proceed to show.

The case r=1 is the easiest to verify. If $|g'|\leq M$, a.e. and $h(\theta) = g(\cos\theta)$, then $|h'(\theta)| = |\sin\theta||g'(\cos\theta)| \leq M|\sin\theta|$, a.e.. From this one easily checks that

$$|\Delta_t^2(h,\theta)| \leq |h(\theta+2t)-h(\theta+t)| + |h(\theta+t)-h(\theta)|$$
$$\leq 3M(|\sin\theta||t| + t^2).$$

This gives that

$$|h(\theta)-L_n(h,\theta)| \leq \int_{-\pi}^{\pi} |\Delta_t^2(h,\theta)|K_n(t)dt$$
$$\leq 3M\int_{-\pi}^{\pi} (|t||\sin\theta|+t^2)K_n(t)dt \leq CM\{|\sin\theta|n^{-1} + n^{-2}\}$$

because of (7.3). When this is written in terms of Λ_n, we get (7.2).

The essential property that allowed us to give the pointwise estimate above was that the appearance of the $\sin\theta$ in h' could be turned into a profit. The following lemma is the same.

LEMMA 7.1. *Let* L_n *be defined as in* (7.4) *with* K_n *satisfying* (7.3). *If* $h \in C(2\pi)$ *with* $|h^{(\nu)}(\theta)| \le A\,|\sin\theta|^{\mu}$, a.e., *and* $0 \le \mu \le r$, $\mu+\nu \le 2r$, *then*

$$|L_n(h,\theta)-h(\theta)| \le CAn^{-\nu}(|\sin\theta|^{\mu} + n^{-\mu}), \quad -\pi \le \theta \le \pi,$$

with C *a constant depending only on* r.

Proof. The proof is similar to that given above for r=1. See [15] for details. Similar arguments are also given in [22].

Any function $h(\theta) = g(\cos\theta)$ with $g \in W_{r,\infty}(I)$ can be decomposed into functions which satisfy Lemma 7.1.

LEMMA 7.2. *Let* L_n *be as in* (7.4). *If* $g \in W_{r,\infty}(I)$ *with* $\|g^{(i)}\|_\infty \le M$, $i=0,1,\dots,r$ *and* $h(\theta) = g(\cos\theta)$, *then*

$$|h(\theta)-L_n(h,\theta)| \le CM(\Delta_n(\cos\theta))^r, \quad -\pi \le \theta \le \pi, \tag{7.6}$$

with C *depending only on* r.

Proof. It is enough to consider the case when M = 1. We can decompose h into a sum of functions which satisfy the hypotheses of Lemma 7.1.

The decomposition goes as follows. Differentiate h r times and write $h^{(r)} = h_o^{(r)} + R_o^{(r)}$, where $h_o^{(r)}(\theta) = (-\sin\theta)^r g^{(r)}(\cos\theta)$ and $R_o^{(r)}$ the collection of the remaining terms. The functions h_o and R_o are defined as the r fold mean value zero integrals of $h_o^{(r)}$ and $R_o^{(r)}$, respectively. This convention holds throughout.

Now, $R_o^{(r)}$ has only terms involving $g^{(i)}$ with $i<r$ and thus has another derivative. Write $R_o^{(r+1)} = h_1^{(r+1)} + R_1^{(r+1)}$, with

$h_1^{(r+1)}$ the collection of terms involving $g^{(r)}$ and $R_1^{(r+1)}$ all the remaining terms. We continue in this way. At the k-th stage, we have $R_{k-1}^{(r+k)} = h_k^{(r+k)} + R_k^{(r+k)}$, with $h_k^{(r+k)}$ all the terms involving $g^{(r)}$ and $R_k^{(r+k)}$ the rest of the terms. We stop at $k=r-1$ and obtain the decomposition $h = h_o+\cdots+h_r$, where h_r is defined to be $R_{r-1}+c$, with c the mean value of h.

It is easy to verify that each h_k, $k=0,\ldots,r$ satisfies Lemma 7.1 with $\mu=r-k$ and $\nu=r+k$ and A a constant depending only on r, e.g. $A=(4r)^{4r}$ surely suffices. Thus, from Lemma 7.1 we find

$$|h(\theta)-L_n(h,\theta)| \leq CA\sum_{k=o}^{r}(|\sin\theta|^k n^{-2r+k}+n^{-2r}) \leq C(\Delta_n(\cos\theta))^r$$

for $-\pi\leq\theta\leq\pi$. Since C depends only on r, the lemma is proved.

THEOREM 7.1. _Let_ L_n _be as in_ (7.4). _Define_ Λ_n _as in_ (7.5). _If_ $f\epsilon C(I)$, _then_ $\Lambda_n(f)$ _is an algebraic polynomial of degree_ $\leq n$ _such that for any_ $n\geq r$,

$$|f(x)-\Lambda_n(f,x)| \leq C\omega_r(f,\Delta_n(x)), \quad -1\leq x\leq 1,$$

with C _a constant depending only on_ r.

Proof. Consider first a function $g\epsilon W_{r,\infty}(I)$. The function $g_1 = g-P_r(g)$ satisfies the hypotheses of Lemma 7.2, with $M = 2^r\,\|g^{(r)}\|_\infty$. This is shown by using the fact that $g_1^{(r)} = g^{(r)}$ and $g_1^{(i)}$ has at least one zero in I, for each $i=0,\ldots,r-1$ because of Rolle's theorem. Hence,

$$\|g_1^{(i)}\|_\infty \leq 2\,\|g_1^{(i+1)}\|_\infty \leq \cdots \leq 2^{r-i}\,\|g_1^{(r)}\|_\infty = 2^{r-i}\,\|g^{(r)}\|_\infty.$$

Now making the substitution $x = \cos\theta$ and using Lemma 7.2 we have

$$(7.7)\quad |g(x)-\Lambda_n(g,x)| \leq C\,\|g^{(r)}\|_\infty(\Delta_n(x))^r, \quad -1\leq x\leq 1,$$

with C depending only on r. This is our estimate for functions $g\epsilon W_{r,\infty}(I)$. Since the operators Λ_n are uniformly bound-

ed, the theorem follows from Theorem 2.4.

Let us now turn to another intriguing question concerning pointwise approximation by algebraic polynomials. That is, whether it is possible to drop the n^{-2} term that appears in the definition of $\Delta_n(x)$ and still have an estimate like that in Theorem 7.1. This question was posed be G.G. Lorentz at the Oberwolfach conference on approximation theory and also by S.B. Steckin in the Soviet Union. Subsequently S. Teljakovskii [31] showed this was true for $r=1$ and I. Gopengauz [20] gave this for $r=2$. We want to show how the operators Λ_n given above can be used to give the case $r=2$, and this of course contains the case $r=1$.

Starting with the operators Λ_n of Theorem 7.1 with $r=2$, if $f\in C(I)$, define

(7.8) $\mu_n(f,x) = \Lambda_n(f,x) + \ell_n(f,x)$,

where $\ell_n(f,x) = -\frac{1}{2}(x-1)(f(-1)-\Lambda_n(f,-1)) + \frac{1}{2}(x+1)(f(1)-\Lambda_n(f,1))$.
$\ell_n(f)$ is the linear function which interpolates $f-\Lambda_n(f)$ at the end points -1 and 1. In this way $\mu_n(f)$ is an algebraic polynomial of degree $\leq n$ which interpolates f at -1 and 1.

In order to establish the finer pointwise estimates, it is necessary to see how well $(\Lambda_n(f))'$ approximates f'.

LEMMA 7.3. Let Λ_n be the operators of Theorem 7.1 with $r=2$. If $Q_n=\Lambda_n(g)$, with $g\in C(I)$, $|g''| \leq M$, a.e. on I. Then,

(7.9) $|g'(x)-Q_n'(x)| \leq CM(\Delta_n(x)), \quad -1\leq x\leq 1,$

with C an absolute constant.

Proof. This proof is similar to that of Lemma 7.2 (see [15]). Using arguments similar to that in Lemma 7.2, one proves that for the trigonometric polynomial $T_n(\theta) = Q_n(\cos\theta)$ and $h(\theta) = g(\cos\theta)$, we have

(7.10) $|h^{(i)}(\theta)-T_n^{(i)}(\theta)| \leq CMn^i(\Delta_n(\cos\theta))^2, \quad i=0,1,2, \ -\pi\leq\theta\leq\pi.$

This estimate can be improved when i=1 and θ is near 0 because then $h-T_{n-1}$ is an even function and hence $h'(0)-T_n'(0) = 0$. When $|\sin \xi| \leq n^{-1}$, we have from (7.10) that $|h''(\xi)-T_n''(\xi)| \leq CMn^{-2}$. Hence, using the mean value theorem, we have for $\theta\varepsilon [-\pi/2,\pi/2]$ and $|\sin \theta| \leq n^{-1}$, that

$$(7.11) \qquad |h'(\theta)-T_n'(\theta)| \leq |h''(\xi)-T_n''(\xi)||\theta| \leq CMn^{-2}|\sin\theta|.$$

This inequality also holds when $\theta \notin [-\pi/2,\pi/2]$ provided that $|\sin \theta| \leq n^{-1}$, because $h'(\pm\pi)-T_n'(\pm\pi) = 0$.

If we superimpose (7.11) and (7.10) for i=1, we have

$$|h'(\theta)-T_n'(\theta)| \leq CM|\sin\theta|\Delta_n(\cos \theta), \qquad \pi\leq\theta\leq\pi.$$

This last inequality put in terms of g' and Q_n' gives (7.9).

THEOREM 7.2. Let Λ_n be as in Theorem 7.1 for r=2. Define μ_n as in (7.8). Then for each $f\varepsilon C(I)$, we have $\mu_n(f)$ is an algebraic polynomial of degree $\leq$ n for which

$$(7.12) \qquad |f(x)-\mu_n(f,x)| \leq C\omega_2(f,n^{-1}(1-x^2)^{\frac{1}{2}}), \qquad -1\leq x\leq 1, \qquad n\geq 2,$$

with C an absolute constant.

Proof. Let $g\varepsilon W_{2,\infty}(I)$ with $\|g''\|_\infty = M$. For such g, we have from Theorem 7.1 that $|g(\pm 1)-\Lambda_n(g,\pm 1)| \leq CMn^{-4}$ and hence $\|\ell_n(g)\|_\infty \leq CMn^{-4}$. From Theorem 7.1, we have

$$(7.13) \qquad |g(x)-\mu_n(g,x)| \leq |g(x)-\Lambda_n(g,x)| + \|\ell_n(g)\|_\infty$$
$$\leq CM((\Delta_n(x))^2 + n^{-4}) \leq CM(\Delta_n(x))^2, \qquad -1\leq x\leq 1.$$

This inequality can be improved near the end points by using the fact that $\mu_n(g,\pm 1) = g(\pm 1)$. For example, when $0\leq x\leq 1$,

$$(7.14) \qquad |g(x)-\mu_n(g,x)| \leq |x-1||g'(\xi)-(\mu_n(g))'(\xi)|$$
$$\leq |x-1| \left(|g'(\xi)-(\Lambda_n(g))'(\xi)| + CMn^{-4}\right)$$
$$\leq CM|x-1|(\Delta_n(\xi) + n^{-4}) \leq CM(1-x^2)(\Delta_n(x)),$$

where in the first inequality we used the mean value theorem with $x<\xi<1$, in the second inequality we used the fact that

$\|(\ell_n(g))'\|_\infty \le CMn^{-4}$, in the third inequality we used Lemma 7.3 and in the last inequality we used the fact that $1-x \le 1-x^2$, for $0\le x\le 1$, and $\Delta_n(\xi) \le \Delta_n(x)$, because $0\le x<\xi$. The same inequality as (7.14) holds when $-1\le x\le 0$.

When we superimpose the two inequalities (7.13) and (7.14) we have

$$(7.15)\quad |g(x)-\mu_n(g,x)| \le CM(1-x^2)n^{-2}, \quad -1\le x\le 1,$$

with C an absolute constant. This is our inequality for functions in $W_{2,\infty}(I)$. The theorem now follows from Theorem 2.4 since the operators μ_n are uniformly bounded.

8 Monotone approximation

Recently, there has been more attention given to approximation with constraints. The appearance of a constraint can make it more difficult to obtain Jackson type estimates. Constraints with a fixed, finite number of functional equations or inequalities usually pose no difficulties once the degree is large enough. However, when the constraint involves an infinite number of inequalities, that are perhaps incompatible with the norm, then the correct direct estimates may be quite formidable to obtain. Still, the general lines of attack developed for the non-constrained problem can be very useful. We want to indicate what modifications are necessary to push this approach through. We do this for monotone approximation, which has been a prototype for constrained problems.

In monotone approximation, we are given a monotone nondecreasing function f ($f\uparrow$) on I and we want to approximate f in turn by polynomials or splines that are likewise nondecreasing. The question then is: does the constraint cost us anything or can we achieve the same degree of approximation as in the nonconstrained case?

It is again meaningful to break up the problem into two

subproblems: i) approximate f by a smooth function g, ii) approximate g by monotone polynomials or splines as the case may be. Now, it is clear that in i), we should not approximate by arbitrary g's but we should try to retain the monotonicity. This leads then to the following definition of a constrained K-functional.

$$(8.1) \quad K_r^{\uparrow}(f,t) = \inf_{g\in W_{r,\infty}(I),\ g\uparrow} \{\|f-g\|_\infty + t\,\|g^{(r)}\|_\infty\}.$$

In Section 2, we were able to show that $K_r(f,t^r)$ and $\omega_r(f,t)$ are asymptotically equivalent for each $f\in C(I)$. But the constructions given there do not preserve monotonicity. Therefore, we immediately have the question of how $K_r^{\uparrow}$ compares to ω_r. For r=1, it is easy to show that $K_r^{\uparrow}(f,t)$ and $\omega_r(f,t)$ are equivalent when $f\uparrow$ by using spline approximation.

Let $\Pi_n = \{x_i^{(n)}\}_o^{2n}$, with $x_i^{(n)} = -1+in^{-1}$ $i=0,1,\dots,2n$. When r=1, the piecewise linear spline $S_n\in\mathcal{S}_2(\Pi_n)$ which interpolates f at the knots will be monotone nondecreasing and satisfy

$$(8.2) \quad \|f-S_n\|_\infty \le \omega_1(f,n^{-1}), \quad n=1,2,\dots$$

$$(8.3) \quad \|S_n'\|_\infty \le n\omega_1(f,n^{-1}), \quad n=1,2,\dots .$$

If t>0, then take n so that $(n+1)^{-1} \le t \le n^{-1}$. Then,

$$(8.4) \quad K_r^{\uparrow}(f,t) \le \|f-S_n\|_\infty + t\,\|S_n'\|_\infty \le 2\omega_1(f,n^{-1}) \le 4\omega_1(f,t).$$

We already know the reverse inequality because of Theorem 2.1.

For r=2, we can use the variation diminishing splines of I. Schoenberg. Let Π_n be as above and let $V_{n,m}(f)$ be the variation diminishing spline of order m with knots Π_n (see [26] for the definition and properties of variation diminishing splines). If $f\uparrow$, then $V_{n,m}(f)$ is also nondecreasing. Also, it is know (see e.g. [11,Ch.2]), that when $g\in W_{2,\infty}(I)$, then

$$(8.5) \quad \|g-V_{n,m}(g)\|_\infty \le Cn^{-2}\,\|g''\|_\infty,$$

with C a constant depending only on m. Hence, from Theorem 2.3

we have that for an arbitrary $f \in C(I)$,

(8.6) $\|f - V_{n,m}(f)\|_\infty \leq C\omega_2(f, n^{-1})$, $n=1,2,\ldots,$

with C again depending only on m.

For our purposes, we need only take $m=3$ and let $S_n = V_{n,3}(f)$. Then, it is easy to estimate S_n''. On each interval $(x_i^{(n)}, x_{i+1}^{(n)})$, $0 \leq i < 2n$, $S_n'' = \alpha_i$ is a constant. When $0 < 2t < n^{-1}$,

$$|\alpha_i| = |t^{-2}\Delta_t^2(S_n, x_i^{(n)})| \leq t^{-2}\{|\Delta_t^2(f - S_n, x_i^{(n)})| + |\Delta_t^2(f, x_i^{(n)})|\} \leq Ct^{-2}\{\omega_2(f, n^{-1}) + \omega_2(f,t)\},$$

where we used (8.6). Now, take $t = 1/3n$ to find

(8.7) $\|S_n''\|_\infty = \max_i |\alpha_i| \leq Cn^2\, \omega_2(f, n^{-1})$, $n=1,2,\ldots,$

with C an absolute constant. When (8.6) and (8.7) are used as in the proof of (8.4), we get

(8.8) $K_2^{\uparrow}(f, t^2) \leq C\omega_2(f,t)$, $t>0$,

with C an absolute constant. The converse of (8.8) is given in Theorem 2.1.

It is still not known whether (8.4) and (8.8) extend to arbitrary r. However, a weaker estimate for general r can be obtained using the following theorem on spline approximation.

THEOREM 8.1. *Let* $\Pi_n = \{x_i^{(n)}\}_0^{2n}$, $x_i^{(n)} = -1 + in^{-1}$, $0 \leq i \leq 2n$. *If* $f \in C(I)$, $f\uparrow$, *and* $r=1$ *or* 2, *then there are splines* $S_n \in \mathcal{S}_r(\Pi_n)$, *with* $S_n\uparrow$, *and*

(8.9) $\|f - S_n\|_\infty \leq C\omega_r(f, n^{-1})$, $n=1,2,\ldots,$

with C *an absolute constant. For* $r>2$, *and* $f^{(k)} \in C(I)$, $k<r$, $f\uparrow$ *there are splines* $S_n \in \mathcal{S}_r(\Pi_n)$, *with* $S_n\uparrow$, *and*

(8.10) $\|f - S_n\|_\infty \leq Cn^{-k}\omega_1(f^{(k)}, n^{-1})$, $n=1,2,\ldots,$

with C *a constant depending only on* r.

Proof. The proof of the existence of splines which satisfy

(8.9) was already given above. The proof of (8.10) is much more difficult and is given in [13]. The construction of these splines is non-linear and highly local. We dont want to give the details but nevertheless, we can indicate briefly why such a construction is plausible.

Let $\varepsilon_n = n^{-k}\omega_1(f^{(k)},n^{-1})$. If $f^{(k)}\varepsilon C(I)$, with $k\geq 2$ (we already know the cases k=0,1 from (8.6)), then the spline $S_n = L_{\Pi_n,\infty}$ given in (4.10) will satisfy

(8.11) $\|f^{(i)}-S_n^{(i)}\|_\infty \leq Cn^i\varepsilon_n$, $i=0,1$.

This estimate for i=0 is already proved in Theorem 4.2 and the estimate for i=1 is proved similarily (see [4]). Hence, if $f\uparrow$ and $f'(x)\geq Cn\varepsilon_n$, $x\varepsilon I$, then by virtue of (8.11) with i=1, the spline S_n will have $S_n'(x)\geq 0$, $x\varepsilon I$, and so $S_n\uparrow$, as desired.

On the other hand, when $f'(x)\leq Cn\varepsilon_n$, x I, then as shown in (8.6), the spline $S_n = V_{n,r}(f)$ satisfies

$$\|f-S_n\|_\infty \leq Cn^{-1}\|f'\|_\infty \leq C\varepsilon_n.$$

We know that $S_n\uparrow$ and so again we have the desired result.

Therefore, we know the validity of (8.10) when f' is always bigger than $Cn\varepsilon_n$ on I and also when f' is always smaller than $Cn\varepsilon_n$ on I. What is needed then is to blend these two results. This is done by breaking up the interval I into intervals where f' is "large" and "small" respectively and then construct appropriate spline approximation for the separate cases.

Theorem 8.1 allows us to give an estimate for $K_r^\uparrow$ in terms of smoothness.

THEOREM 8.2. _If_ $f\varepsilon C(I)$ _and_ $f\uparrow$, _then_ for r=1,2,

(8.12) $K_r^\uparrow(f,t^r) \leq C\omega_r(f,t)$, $t>0$,

with C _an_ _absolute_ _constant_. _For_ r>2, $0\leq k<r$, $f^{(k)}\varepsilon C(I)$, _and_ $f\uparrow$

(8.13) $K_r^\uparrow(f,t^r) \leq Ct^k\omega_1(f,t)$, $t>0$,

with C depending only on r.

Proof. We have already shown (8.12). To prove (8.13), we use the splines of Theorem 8.1. For $r>2$, according to Theorem 8.1, there is a spline $S_n \in \mathcal{S}_{r+1}(\Pi_n)$ such that $S_n \uparrow$ and

(8.14) $\|f-S_n\|_\infty \le Cn^{-k}\omega_1(f^{(k)},n^{-1})$.

Similar to the way we estimated the derivative of $V_{n,3}(f)$ in (8.7), it can be shown (see [14] for details) that

(8.15) $\|S_n^{(r)}\|_\infty \le Cn^{r-k}\omega_1(f^{(k)},n^{-1})$,

with C depending only on r. For $0<t$, we choose n so that $(n+1)^{-1}<t\le n^{-1}$ and then use S_n in the same way we estimated in (8.4) to obtain (8.13).

Monotone approximation by algebraic polynomials parallels to a large extent that of spline approximation in that it is relatively easy to obtain the Jackson estimates in terms of ω_1 and ω_2, but beyond that things become much more difficult. Of course, now we have the advantage of using Theorem 8.2, which was proved using spline approximation.

The lower order estimates for monotone polynomial approximation follow from the constructions of G. G. Lorentz - K. Zeller [24], $r=1$, and G. G. Lorentz [23], $r=2$. The Lorentz-Zeller proof is to go to the trigonometric case via the substitution $x=\cos\theta$. The function $h(\theta) = f(\cos\theta)$ is now a bell shaped function which in turn must be approximated by bell shaped trigonometric polynomials of degree n. This is done by approximating h first by an interpolating spline of order r with 2n equally spaced knots and then using the Jackson operators of Section 5 to approximate the spline by a trigonometric polynomial.

It is also possible to argue directly in the same spirit as Section 6. Consider the case $r=2$ which includes the case $r=1$, since $\omega_2(f,t) \le 2\omega_1(f,t)$, $t>0$. Suppose that $g\in W_{2,\infty}(I)$,

$\|g''\|_\infty = M$ and $g\uparrow$. First extend g outside of I by setting $g(x) = g(1)+g'(1)(x-1)$, $x>1$ and $g(x) = g(-1)+g'(-1)(x+1)$, $x<-1$. The extended g is still nondecreasing and has $\|g''\|_\infty[-2,2] = M$. The fact that we can extend g monotonically without losing smoothness is one of the basic steps that can not be done in general.

We can also make a linear adjustment without changing the monotonicity of g and thereby assume that $g(-2)=0$ and $g'(a)=0$, for some $-2<a<2$. Analogous to Section 6, we set

$$(8.16)\quad L_n(g,x) = \int_{-2}^{2} g(t)\mu_n(x-t)dt,$$

with for each n, μ_n a non-negative, even algebraic polynomial of degree n with integral 1 over $[-2,2]$ and

$$(8.\quad \int_{-3}^{3} t^2\mu_n(t)dt \le Cn^{-2}, \quad n=1,2,\ldots,$$

with C an absolute constant. We also assume that

$$(8.17)\quad \|\mu_n\|_\infty[-3,-1] \le C_1 n^{-2}.$$

The polynomials μ_n of Section 6 satisfy all the above.

As we argued in Section 6,(see [11]), one can show that

$$\|g-L_n(g)\|_\infty \le CMn^{-2},$$

with C an absolute constant. The polynomials $L_n(g)$ are not necessarily nondecreasing, but we have for $x\varepsilon I$,

$$(L_n(g))'(x) = \int_{-2}^{2} g'(t)\mu_n(x-t) + g(-2)\mu_n(x+2) - g(2)\mu_n(x-2)$$
$$\ge -g(2)C_1 n^{-2}$$

because g' and μ_n are non-negative, $g(-2)=0$ and (8.17).

Thus, the polynomials $P_n(x) = L_n(g,x) + C_1 g(2)n^{-2}x$ will be *nondecreasing and satisfy*

$$(8.18)\quad \|g-P_n\|_\infty \le \|g-L_n(g)\|_\infty + C_1 n^{-2}\|g\|_\infty \le CMn^{-2},$$

with C an absolute constant. Here, we used the fact that $\|g\|_\infty \le 16\,\|g''\|_\infty$, which follows because both g and g' have a

zero in $[-2,2]$. This is our result for smooth functions and leads to the following theorem.

THEOREM 8.3. <u>If</u> $f \in C(I)$, $f\uparrow$, <u>and</u> $r=1$ or 2, <u>then for each</u> $n \geq 1$, <u>there is a polynomial</u> P_n <u>of degree</u> $\leq n$, <u>such that</u> $P_n \uparrow$ <u>and</u>

$$(8.19)\quad \|f-P_n\|_\infty \leq C\omega_r(f,n^{-1}),$$

<u>with</u> C <u>an absolute constant</u>. <u>For</u> $r>2$, $f^{(k)} \in C(I)$, $0 \leq k<r$, $f\uparrow$, <u>we have that for each</u> $n \geq 1$, <u>there is a polynomial</u> P_n <u>of degree</u> $\leq n$, <u>such that</u> $P_n \uparrow$ <u>and</u>

$$(8.20)\quad \|f-P_n\|_\infty \leq Cn^{-k}\omega_1(f^{(k)},n^{-1}),$$

<u>with</u> C <u>a constant depending only on</u> r.

<u>Proof</u>. For $r=2$ and $f \in C(I)$, take $g\uparrow$ so that $\|f-g\|_\infty + n^{-2}\|g''\|_\infty \leq C\omega_2(f,n^{-1})$. The existence of g is guaranteed by Theorem 8.2. Let $P_n\uparrow$ be the polynomial which approximates g according to (8.18). Then,

$$\begin{aligned}\|f-P_n\|_\infty &\leq \|f-g\|_\infty + \|g-P_n\|_\infty \leq \|f-g\|_\infty + Cn^{-2}\|g''\|_\infty \\ &\leq C\omega_2(f,n^{-1}),\end{aligned}$$

with C an absolute constant. This is (8.19) for $r=2$ which also also includes the case $r=1$.

For $r>2$, the story is much the same as for splines. The proof is detailed and is given in [14]. While the matter is a little simplified because of Theorem 8.2, we no longer have the nice B-splines with which to make local corrections. This requires the construction of polynomials which mimic the B-splines in that they are large on a given interval and fall off fast outside of this interval.

<u>References</u>

1 Bak, J. and D.J. Newman, Müntz-Jackson theorems in $L^p[0,1]$ and $C[0,1]$, American J. of Math., <u>154</u> (1972), 437-457.

2 Berens, H., <u>Interpolationsmethoden zur Behandlung von Approximationsprozessen auf Banachraümen</u>, Springer Lecture

Notes in Mathematics, no. 64, Springer, Berlin, 1968.

3 Besov, O.V., Continuation of functions with conservation of their differential and difference properties in L_p, Soviet Math. Dokl., 4 (1963), 770-774.

4 deBoor, C. and G. Fix, Spline approximation by quasi-interpolants, J. Approximation Theory, 8 (1973), 19-45.

5 Brudnyi, Ju.A., Generalizations of a theorem of A.F. Timan, Soviet Math. Dokl., 4 (1963), 244-247.

6 Brudnyi, Ju.A., Approximation of functions by algebraic polynomials, Math. of USSR-Isv., 2 (1968), 735-743.

7 Butzer, P.L. and H. Berens, Semi-Groups of Operators and Approximation, Springer, Berlin, 1967.

8 Butzer, P.L. and K. Scherer, Approximationsprozesse und Interpolationsmethoden, Bibliographisches Institut AG, Mannheim, 1968.

9 Curry, H.B. and I.J. Schoenberg, On Polya frequency functions IV, the fundemental splines and their limits, J. d'Analyse Math., 17 (1966), 71-107.

10 DeVore, R., On Jackson's theorem, J. Approximation Theory, 1 (1968), 314-318.

11 DeVore, R., Approximation of Continuous Functions by Positive Linear Operators, Springer Lecture Notes in Mathematics, no. 293, Springer, Berlin, 1972.

12 DeVore, R., Degree of monotone approximation, in Linear Operators and Approximation, II, P.L. Butzer and B. Sz. Nagy ed., Birkhäuser Verlag, Basel, 1974, pp. 337-352.

13 DeVore, R., Monotone approximation by splines, SIAM Math. Anal., to appear.

14 DeVore, R., Monotone approximation by polynomials, SIAM Math. Anal., to appear.

15 DeVore, R., Pointwise approximation by polynomials and splines, to appear in the proceedings of the conference on the constructive theory of functions, Kalouga, 1976.

16 Dzadyk, V.K., A further strengthening of Jackson's theorem on the approximation of continuous functions by ordinary polynomials, Doklady, 121 (1958), 403-406.

17 Freud, G., Sui procedimenti lineari d'approssimazione, Rend. Mat.,(8) 26 (1959), 641-643.

18 Freud, G., Über die Approximation reeler stetiger Funktionen durch gewöhnliche Polynome, Math. Anal., 137 (1959),

17-25.

19 Freud, G. and V. Popov, On approximation by spline functions, Proc. Conf. on Constructive Theory of Functions, G. Alexits and S.B. Stechkin ed., Akademia Kiado, Budapest, 1969, pp.163-172.

20 Gopengauz, I.E., A question concerning the approximation of functions on a segment and a region with corners, Teor. Funkcii Funkcional. Anal. i Pril.. 4 (1967), 204-210.

21 Johnen, H., Inequalities connected with the moduli of smoothness, Math. Vestnik, 9(24), 289-303.

22 Lorentz, G.G., Approximation of Functions, Holt, N.Y., 1966.

23 Lorentz, G.G., Monotone approximation, in Inequalities,III, O. Shisha ed., Academic Press, N.Y., 1972, pp. 201-215.

24 Lorentz, G.G. and K.L. Zeller, Degree of approximation by monotone polynomials II, J. Approximation Theory 2 (1969), 265-269.

25 Lyche T. and L.L. Schumaker, Local spline approximation methods, J. Approximation Theory, to appear.

26. Marsden, M., An identity for spline functions with applications to variation diminishing spline approximation, J. Approximation Theory, 3 (1970), 7-49.

27 Peetre, J., A Theory of Interpolation of Normed Spaces, Lecture Notes, Brazilia, 1963.

28 Peetre, J., On the connections between the theory of interpolation spaces and approximation theory, in Proc. Conf. on Constructive Theory of Functions, G. Alexits and S.B. Stechkin ed., Akademia Kiado, Budapest, 1969, pp.354-363.

29 Popov V.A. and Bl.K. Sendov, The classes that are characterized by the best approximation by spline functions, Math. Notes 8 (1970), 550-557.

30 Scherer, K., Über die beste Approximation von L^p-Funktionen durch Splines, in Proc. of Conference on Constructive Function Theory, Varna, 1970, pp. 277-286.

31 Teljakovskii, S.A., Two theorems on the approximation of functions by algebraic polynomials, Mat. Sbornik, 70(112) (1966), 252-265.

32 Timan, A.F., Strengthening of Jackson's theorem on the best approximation of continuous functions on a finite segment of the real axis, Doklady 78 (1951), 17-20.

33 Whitney, H., Analytic extensions of differentiable functions defined in closed sets, Trans. Amer. Math. Soc.,

36 (1934), 63-89.

34 Whitney, H., On functions with n-th bounded differences, J. de Math. pure et appl. 36(1957), 60-95.

35 Varga, R., Error bounds for spline interpolation, in Approximations with Special Emphasis on Spline Functions, I.J. Shoenberg ed., Academic Press, N.Y., 1969, pp. 367-388.

36 Zygmund A., Smooth functions, Duke Math. J., 12 (1945), 47-76.

Ronald A. DeVore +
Department of Mathematics
Oakland University
Rochester, MI 48063
and
Mathematisches Institut
Universität Erlangen-Nürnberg
8520 Erlangen, W. Germany

+ Supported in part by the National Science Foundation Grant GP 19620 and the Alexander von Humboldt-Stiftung.

LAGRANGE INTERPOLATION AT ZEROS OF ORTHOGONAL POLYNOMIALS

G. P. Nevai

The purpose of this paper is to give a detailed survey of new results in the theory of convergence of Lagrange interpolation taken at zeros of orthogonal polynomials with special emphasis on results related to infinite intervals. Many open problems are indicated.

When Professor G. G. Lorentz asked me to write a short survey I decided to choose the title "Interpolation at zeros of orthogonal polynomials and related problems." What should a survey with such a title have contained? It should certainly have dealt with problems of convergence and divergence of Lagrange, Hermite-Fejér, lacunary and trigonometric interpolation, as well as with those of mechanical quadrature processes. It should have discussed some aspects of the general theory of orthogonal polynomials, in particular asymptotic relations and bounds for orthogonal polynomials and the distribution of zeros of orthogonal polynomials. Questions of weighted uniform and L^p approximation of functions by polynomials should have been mentioned also. Furthermore, I should have considered problems of stability of interpolation and that of "rough and fine" interpolation theory. The use of interpolation in solving other problems should also have been discussed. In short, I soon realized that such an article would be impossibly long. Thus I decided to restrict my attention to convergence and divergence of Lagrange interpolation, this being the topic with which I am most familiar.

This survey will be in no sense complete. For instance, neither Fejér's nor Szegö's very important results will be

Preceding page blank

mentioned. I will concentrate only on recent results, but even so, the paper will be far from exhaustive. The material considered is divided into two parts; the first deals with pointwise convergence and the second with mean convergence.

Let α be a bounded, non-decreasing function defined on $\mathbb{R}$. We say that α is a weight function if it has infinitely many points of increase and all the moments of α are finite. If α is absolutely continuous we write $w = \alpha'$. For a given weight α there exists a unique sequence of polynomials $p_n(d\alpha,x) = \gamma_n(d\alpha)x^n + \ldots$ $(n = 0,1,\ldots)$ having the properties $\gamma_n(d\alpha) > 0$ and

$$\int_{-\infty}^{\infty} p_n(d\alpha,x)p_m(d\alpha,x)d\alpha(x) = \delta_{nm}.$$

The zeros of $p_n(d\alpha)$ are real and distinct; we denote them by $x_{kn}(d\alpha)$: $x_{1n}(d\alpha) > x_{2n}(d\alpha) > \ldots > x_{nn}(d\alpha)$. For a given function f the Lagrange interpolation polynomial $L_n(d\alpha,f)$ corresponding to the weight α is defined to be the unique algebraic polynomial of degree at most $n-1$ which coincides with f at the nodes $x_{kn}(d\alpha)$ $(k = 1,2,\ldots,n)$. Thus

$$L_n(d\alpha,f,x) = \sum_{k=1}^{n} f(x_{kn}(d\alpha))\ell_{kn}(d\alpha,x)$$

where $\ell_{kn}(d\alpha)$ are the fundamental polynomials. An important quantity we shall need later is the Christoffel function $\lambda_n(d\alpha)$ corresponding to the weight α. It may be defined in several ways; for instance,

$$\lambda_n(d\alpha,x) = \min_{\substack{\pi\in\mathbb{P}_{n-1}\\ \pi^2(x)=1}} \int_{-\infty}^{\infty} \pi^2(t)d\alpha(t)$$

where $\mathbb{P}_n$ denotes the set of polynomials of degree at most n.

By $w^{(a,b)}$ we shall denote the Jacobi weight, that is, $w^{(a,b)}(x) = (1-x)^a(1+x)^b$ for $|x| < 1$ and $w^{(a,b)}(x) = 0$; otherwise, $(a,b > -1)$.

I. Pointwise convergence

1. We begin with the classical and most thoroughly investigated part of the theory of Lagrange interpolation, namely, pointwise convergence of $L_n(d\alpha, f)$ when the support of $d\alpha$ is compact. Recent investigations here may be classified as follows:

(i) Consideration of $L_n(d\alpha, f)$ when α is not necessarily absolutely continuous.

(ii) Finding estimates which hold uniformly on the whole support of $d\alpha$ for the deviation of $L_n(d\alpha, f)$ from f and similarly for the corresponding Lebesgue function.

(iii) Improvement of convergence either by adding a few additional points to the set of nodes or by using certain other tricks.

(iv) Determination of conditions assuring divergence.

Because of the huge number of results, we will illustrate the above classification with only a few selected theorems. Concerning (i), G. Freud [28] and Ja. L. Geronimus [49] have obtained many interesting theorems, among which the following seems to be the strongest.

THEOREM 1. *Let* $\operatorname{supp}(d\alpha)$ *be compact and let* $|p_n(d\alpha,x)|$ *be uniformly bounded for* $x \in m \subset \operatorname{supp}(d\alpha)$. *Let* f *be a function vanishing in a neighbourhood of the interval* Δ *with*

$$(1) \quad \int_{-\infty}^{\infty} |f(t)|^2 d\alpha(t) < \infty,$$

where the integral is understood in Riemann-Stieltjes sense. Then $L_n(d\alpha, f, x)$ *converges to* 0 *uniformly in* $x \in m \cap \Delta$ *when* $n \to \infty$.

Theorem 1 combined with the classical machinery can be used to prove a lot of convergence theorems at individual points. One of them is the following

THEOREM 2. <u>Let</u> x_0 <u>be fixed and let</u> f <u>satisfy</u> (1). <u>Let</u> α <u>be absolutely continuous in a neighbourhood</u> $V(x_0)$ <u>of</u> x_0. <u>If</u> $\alpha'(x) \leq M < \infty$ <u>and</u> $|p_n(d\alpha,x)| \leq M < \infty$ $(n = 1,2,\dots)$ <u>for</u> $x \in V(x_0)$ <u>and</u> f <u>is of bounded variation in</u> $V(x_0)$ <u>then</u>

$$\lim_{n\to w} L_n(d\alpha,f,x_0) = f(x_0)$$

<u>provided that</u> f <u>is continuous at</u> x_0.

For further results we refer the reader to [25], [49], [59], [60], [79] and [81].

The important role of Lebesgue functions is well known. Nevertheless, for general weights only fairly weak estimates have been obtained. We know that if $\mathrm{supp}(d\alpha) \subset [-1,1]$ and α' is bigger than the Chebyshev weight, then the corresponding Lebesgue function is of order $\sqrt{n}$ uniformly in $[-1,1]$. If α' is strictly positive on $[-1,1]$, then the Lebesgue function is of order $\sqrt{n}$ on every interval $\Delta \subset (-1,1)$, and it is of order n on $[-1,1]$. These are the strongest estimates known. If we impose conditions also on $|p_n(d\alpha)]$ we can obtain better estimates as it was done by G. Freud [25] (see also [49]). It should be possible to determine the exact order of Lebesgue functions of Lagrange interpolation corresponding to weights of the form $w = \emptyset w^{(a,b)}$ where $\emptyset$ satisfies some growth and smoothness conditions. For $\emptyset \equiv 1$, G. I. Natanson [70] proved the next

THEOREM 3. <u>Let</u> $a,b > -\frac{1}{2}$ <u>and let</u> $\ell_n(a,b,x)$ <u>denote the Lebesgue function of Lagrange interpolation taken at the zeros of</u> $p_n(w^{(a,b)})$. <u>Then</u> $|\ell_n(a,b,x) - 1|$ <u>is exactly of order</u>

$$|p_n(w^{(a,b)},x)|(1 + w^{(a/2+1/4,b/2+1/4)}(x)\log n)$$

when $n \to \infty$.

We hope to soon obtain similar estimates when $\phi > 0$ and belongs to $\mathrm{lip}\,\beta$. See also [54], [56], and [94].

As mentioned above, if $\mathrm{supp}(d\alpha) \subset [-1,1]$ and $\alpha'(x) \geq m > 0$ for almost every $x \in [-1,1]$, then it follows that for every $f \in \mathrm{lip}\,\frac{1}{2}$ $L_n(d\alpha,f,x)$ converges to $f(x)$ uniformly on $[a,b] \subset (-1,1)$. G. Freud [30] has asked why this convergence is not uniform on the whole interval $[-1,1]$. One possible explanation is that $\mathrm{supp}(d\alpha) \subset [-1,1]$ implies that $|x_{kn}(d\alpha)| < 1$ (and therefore $L_n(d\alpha,f,\pm 1) \neq f(\pm 1)$) and $|p_n(d\alpha,\pm 1)|$ grows so rapidly that $L_n(d\alpha,f,\pm 1)$ cannot converge to $f(\pm 1)$. For this reason G. Freud suggested adding two new points to the set of nodes, namely ± 1. Let us denote by $L_n^*(d\alpha,f)$ the Lagrange interpolation polynomials taken at the nodes $\{x_{kn}(d\alpha)\} \cup \{\pm 1\}$.

THEOREM 4. *Let* $\mathrm{supp}(d\alpha) \subset [-1,1]$ *and* $\alpha'(x) \geq \sqrt{1-x^2}$ *for almost all* $x \in (-1,1)$. *Suppose also that*

$$\int_{-1}^{1} \frac{d\alpha(t)}{\sqrt{1-t^2}} < \infty .$$

If $f \in \mathrm{lip}\,\frac{1}{2}$ *on* $[-1,1]$ *then* $L_n^*(d\alpha,f)$ *converges to* f *uniformly on* $[-1,1]$ *when* $n \to \infty$.

For further development of this idea see [30], [31].

Concerning estimations from below and divergence of Lagrange interpolation we mention two results which seem to be typical.

THEOREM 5. *Let* v *denote the Chebyshev weight. If* ω *is a modulus of continuity with*

$$\lim_{t \to +0} \frac{\omega(t)}{t\,|\log t|} = 0$$

then there exists a continuous function f *on* $[-1,1]$ *with*

$\omega(f,\delta) \le \omega(\delta)$ *and*

$$\limsup_{n\to\infty} \frac{|L_n(v,f,1)-f(1)|}{\omega(1/n)} > 0.$$

THEOREM 6. *Let* supp $(d\alpha) \subset [-1,1]$ *and let* $\ell_n(d\alpha)$ *denote the corresponding Lebesgue constants. Let* ω *be a modulus of continuity. If*

$$\infty \ge A = \limsup_{n\to\infty} \omega\left(\frac{1}{n^2 \ell_n(d\alpha)}\right) \ell_n(d\alpha) > 0$$

then there exists a continuous f *with* $\omega(f,\delta) \le C\omega(\delta)$ *and*

$$\limsup_{n\to\infty} \max_{|x|\le 1} |f(x) - L_n(d\alpha,f,x)| = A.$$

For other results we refer the reader to works of O. Kis, J. Szabados, P. Vértesi, D. L. Berman. The above theorems were proved in [54] and [58], respectively.

Although much more could be said about the case when supp$(d\alpha)$ is compact, we stop here, for there are other things we would like to mention. Less is known about the behavior of Lagrange interpolation processes corresponding to weights with non-compact support. What is known will be discussed in the next three sections.

2. Here we will give two theorems. They are the only known convergence theorems when supp$(d\alpha)$ is non-compact and α is not one of the classical weights. We would like to hope that the reader, realizing how weak the statements below are, will try to find better results. The significance of the next theorem of G. Freud [27] is that it was the first successful attempt to show convergence of Lagrange interpolation on an infinite interval.

THEOREM 7. *Let* $f \in \text{lip}\,\frac{1}{2}$ *on* $\mathbb{R}$. *If* w *satisfies the*

conditions

$$w(x) > c_1 \exp(-ax^2), \quad \int_{-\infty}^{\infty} \exp(c_2 x^2) w(x)dx < \infty$$

$(c_1, c_2, a > 0)$ then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \exp(-\frac{a}{2}x^2) |f(x) - L_n(w,f,x)| = 0.$$

This theorem can be easily generalized to weights w with

$$w(x) > c_1 \exp(-a|x|^{\beta}), \quad \int_{-\infty}^{\infty} \exp(c_2 |x|^{\beta}) w(x)dx < \infty$$

for $\beta = 4, 6, 8 \ldots$ but it would be really interesting to know what happens if $\beta \in \mathbb{R}^+$. If instead of uniform convergence we are interested only in convergence almost everywhere, then the class of weights in consideration can be extended. The following theorem has been proved in [72].

THEOREM 8. Let $\operatorname{supp}(d\alpha)$ be non-compact. Suppose that there exists a function g which satisfies the following two conditions.

(i) $g^{-1} \in L^2_{d\alpha}$ and the Gauss-Jacobi mechanical quadrature process taken at the zeros of $p_n(d\alpha)$ is convergent for $|g|^{-2}$.

(ii) For every non-negative integer n

$$|x^n g(x)| \leq C\cdot (g_n)^n \qquad (x \in \mathbb{R})$$

where $g_n \geq 1$, $g_n n^{-1} \downarrow 0$ when $n \to \infty$. Suppose also that f is bounded on $\mathbb{R}$ and for some natural integer R the modulus of smoothness of R-th order of f satisfies the condition

$$\sum_{n=1}^{\infty} \omega_R(f, g_n n^{-1})^2 < \infty.$$

<u>Then for</u> $d\alpha$ <u>almost every</u> x, $L_n(d\alpha,f,x)$ <u>converges to</u> $f(x)$ <u>when</u> $n \to \infty$.

This theorem can be applied to the case where $\alpha'(x) = |x|^{\epsilon}\exp(-|x|^{\beta})$ $(\epsilon > -1,\ \beta = 2,4,\ldots)$. In this case we can put $g(x) = \exp(-\frac{1}{4}|x|^{\beta})$ and then $g_n = (4n)^{1/\beta}$.

We conclude this section with a problem which is closely related to Freud's theorem. Let $\alpha'(x) = \exp(-|x|^{\beta})(\beta \in \mathbb{R}^+, x \in \mathbb{R})$. What can be said about the magnitude of the sum

$$\sum_{k=1}^{n} \lambda_{kn}(d\alpha)\exp(c|x_{kn}|^{\beta})$$

$(\lambda_{kn}(d\alpha) \equiv \lambda_n(d\alpha,x_{kn}(d\alpha)))$? We suspect that it is bounded for $c < 1$ and it is of order $n^{1/\beta}$ for $c = 1$. The case $c > 1$ is also interesting (see [73]).

3. Here we will be concerned with pointwise convergence of Lagrange interpolation taken at the Hermite abscissas, that is, $w(x) = \exp(-x^2)$, $x \in \mathbb{R}$. Even if we restrict ourselves to consider the simplest estimates, that is, estimates like

$$|f(x) - L_n(w,f,x)| \leq \text{something}\cdot\omega(f,?),$$

there are two problems which arise immediately. First, there is no hope of finding estimates where the right side is independent of $x \in \mathbb{R}$. Secondly, the classical methods which use best approximating polynomials and Lebesgue function estimates might fail since we have functions defined on the whole real line. The first problem can be eliminated easily. Since in our case $w(x) = \exp(-x^2)$ one may expect a factor of $\exp(\frac{1}{2}x^2)$ to appear on the right side. The second problem may be removed in the following way. We know that $L_n(w,f)$ depends only on those values of f which it takes at the Hermite abscissas and the biggest zero of the n-th Hermite polynomial is less than $\sqrt{2n+1}$. So let us apply the Jackson-Brudny theorem on the

interval $(-\sqrt{2n+1}, \sqrt{2n+1})$ and see what happens. The best result one can hope to obtain would be

$$|f(x) - L_n(w,f,x)| \le C \exp(\tfrac{1}{2}x^2) \log n \, \omega_R\left(f, \frac{1}{\sqrt{n}}\right)$$

(here $\omega_R(f,\delta)$ denotes the R-th modulus of smoothness of f). This was the estimate which G. Freud sought to prove in 1968. To his (and others') surprise he managed to obtain a much stronger result:

THEOREM 9. Let f be uniformly continuous on $\mathbb{R}$. Then

$$|f(x)-L_n(w,f,x)| \le C\left[\log n + \exp(\tfrac{1}{2}x^2)\right] \omega_R\left(f, \frac{1}{\sqrt{n}}\right)$$

for every real x and natural integer R with the constant depending only on R.

Theorem 9 cannot be improved, as is shown by the next two theorems (see [29], [57]).

THEOREM 10. For every modulus of continuity $\omega(\delta)$ there exist a uniformly continuous function F, an increasing sequence of natural integers $\{n_r\}$ and a sequence $\{x_r\}$ such that

$$\omega(F,\delta) \le \omega(\delta)$$

and

$$|F(x_r)-L_{n_r}(w,F,x_r)| \ge C\left[\log n_r + \exp(\tfrac{1}{2}x_r^2)\right] \omega\left(\frac{1}{\sqrt{n_r}}\right)$$

for $r = 1,2,\ldots$.

Let ϕ be continuous, nondecreasing on $\mathbb{R}^+$ with $\phi(0)=0$, $\phi(\delta) > 0$ $(\delta > 0)$ and let $\phi(\delta)\delta^{-R}$ be non-increasing on $\mathbb{R}^+$ for a fixed natural integer R with

$$\lim_{\delta\to+0} \phi(\delta)\delta^{-R} = +\infty .$$

THEOREM 11. *There exist a uniformly continuous function* G *and sequences* $\{n_r\}$ and $\{x_r\}$ *such that*

$$\omega_R(G,\delta) \le \phi(\delta)$$

and

$$|G(x_r) - L_{n_r}(w,G,x_r)| \ge C\left[\log n_r + \exp(\tfrac{1}{2} x_r^2)\right] \phi\left(\frac{1}{\sqrt{n_r}}\right)$$

for $r = 1,2,\ldots$.

When G. Freud drew my attention to Lagrange interpolation on infinite interval, the first thing I wanted to prove was the convergence of $L_n(w,f)$ when f is of bounded variation. I had been trying to apply all the known methods to prove this but none of them worked. Then I remembered that in the theory of trigonometric Fourier series the simple formula

$$S_n(x,f) - f(x) = \frac{1}{2}[S_n(x,f) - S_n(x+\frac{\pi}{n}, f)] + O[\omega(f, \frac{1}{n})]$$

had helped me to find many interesting results and I tried to find a similar representation for the deviation of $L_n(w,f)$ from f. This was done in [79]:

THEOREM 12. *Let* f *be uniformly continuous on* $\mathbb{R}$ *and let* $\epsilon > 0$ *be fixed. Then*

$$\sup_{x\in\mathbb{R}} \exp(-\frac{1}{2}x^2)\,\Big|L_n(w,f,x) - f(x) - \frac{1}{2}\sum_{|x-x_{kn}|<\epsilon}[f(x_{kn})- f(x_{k-1,n})]\ell_{kn}(w,x)\Big| = O\left[\omega\left(f,\frac{\log n}{\sqrt{n}}\right)\right]$$

for $n = 1,2,\ldots$

Let us remark that I suspect this is not the best possible estimate; the right side should perhaps be $O\left[\omega\left(f, \frac{1}{\sqrt{n}}\right)\right]$. With the aid of the above formula I could easily prove convergence

for functions having bounded variation. But this formula did more than simply provide a lemma for proving convergence for functions expressible as the difference of two monotonic functions. It turned out that it implicitly contained many convergence tests.

Let us say that f satisfies the one-sided Dini-Lipschitz condition on the (finite or infinite) interval (a,b) if

$$f(x+\delta) - f(x) \geq - \frac{\nu(\delta)}{|\log \delta|} \qquad (a < x < x+\delta < b)$$

where δ is small, $\nu(\delta) \geq 0$ is nondecreasing with $\lim_{\delta\to+0} \nu(\delta) = 0$. It is obvious that every function of bounded variation is the difference of two functions satisfying the one-sided Dini-Lipschitz condition (with $\nu \equiv 0$, say).

THEOREM 13. *Let* f *be uniformly continuous on* $\mathbb{R}$ *and satisfy the one-sided Dini-Lipschitz condition on the (finite or infinite) interval* (a,b). *If* $0 < \epsilon < (b-a)/2$ *then*

$$\lim_{n\to\infty} \sup_{a+\epsilon < x < b-\epsilon} \exp(-\tfrac{1}{2}x^2)\,|f(x) - L_n(w,f,x)| = 0.$$

This theorem does not give the rate of convergence, but the next convergence test will have this feature. Let ϕ be defined on $\mathbb{R}^+$ and let

$$\phi(0) = 0, \quad \phi(x) \leq \phi(y) \qquad (x < y).$$

Following L. C. Young [98] we say that f has bounded ϕ variation on the (finite or infinite) interval (a,b) if

$$v_\phi(f) = \sup_\tau \sum_i \phi(|f(b_i) - f(a_i)|) < \infty$$

where τ is an arbitrary finite system of finite disjoint intervals $(a_i,b_i) \subset (a,b)$.

THEOREM 14. *Let* f *be uniformly continuous on* $\mathbb{R}$ *and be of*

<u>bounded variation on</u> $(a,b) \subseteq \mathbb{R}$. <u>If</u> $\log \emptyset \in L^1(0,1)$ <u>and</u> $0 < \epsilon < \frac{b-a}{2}$ <u>then</u>

$$\sup_{a+\epsilon < x < b-\epsilon} \exp\left(-\frac{1}{2}x^2\right) |f(x) - L_n(w,f,x)| =$$

$$= O\left[\int_0^{\omega(f,1/\sqrt{n})} \left|\log \frac{v_\emptyset(f)}{\emptyset(\xi)}\right| d\xi + \omega\left(f, \frac{\log n}{\sqrt{n}}\right)\right] .$$

If $\emptyset(\xi) = \xi^p$ $(1 \le p < \infty)$, that is, in case of functions of bounded p variation (see [96]), the integral on the right side is of order

$$\left|\log \omega\left(f, \frac{1}{\sqrt{n}}\right)\right| \omega\left(f, \frac{1}{\sqrt{n}}\right).$$

The reader might notice that all the above theorems require uniform continuity on the whole real axis of functions under consideration, even if we are interested in convergence only at individual points. This strong requirement can be dropped by using the following localization theorem proved in [80]. I would like to emphasize that it is by no means clear that this is definitely the best possible version of the theorem.

Let R denote the set of those functions f defined on the real line which are Riemann integrable on every finite interval and are of order $\exp(\epsilon x^2)$ when $x \to \infty$ where $\epsilon = \epsilon(f)$ is non-negative and less than $1/2$. Since Riemann integrable functions are always bounded we may suppose that for every f in R

$$|f(x)| \le A \exp(\epsilon x^2) \quad (x \in \mathbb{R},\ A = A(f),\ \epsilon = \epsilon(f) < \frac{1}{2}).$$

THEOREM 15. <u>If</u> $f \in R$ <u>and</u> $f \equiv 0$ <u>on</u> $[a-\delta, b+\delta]$ $(\delta > 0)$ <u>then</u>

$$\max_{a \le x \le b} \exp\left(-\frac{1}{2}x^2\right) |L_n(w,x,f)| = o(1)$$

when $n \to \infty$ and $o(1)$ depends only on f and δ but not on the interval (a,b).

The proof of Theorem 15 is rather complicated and is based on two facts. One of them is the definition of the Riemann integral. The other is the inequality

$$\sum_{\substack{|x-x_{kn}|>\delta \\ |x_{kn}|\leq\sqrt{n}}} |\exp(\epsilon x_{kn}^2)\ell_{kn}(w,x) + \exp(\epsilon x_{k+1,n}^2)\ell_{k+1,n}(w,x)| \leq$$

$$\leq C \exp(\tfrac{1}{2} x^2) \frac{1}{\sqrt{n}}$$

for $\delta > 0$, $0 < \epsilon < \frac{1}{2}$, $|x| \leq \sqrt{n}$.

Theorem 15 can be used to prove localized convergence theorems. We shall mention three conditions assuring convergence of $L_n(w,f)$ to f and a given point x_0.

THEOREM 16. Let $f \in R$ and x_0 be fixed. Let f satisfy one of the following three conditions:

(i) $|f(x_0 \pm t) - f(x_0)| \leq \mu(t) \qquad (0 \leq t \leq \delta)$

where μ is a nondecreasing function with

$$\lim_{\rho\to+0} \int_\rho^\delta \frac{\mu(t)}{t}\,dt < \infty.$$

(ii) f is continuous at x_0 and satisfies the one-sided Dini-Lipschitz condition in a neighborhood of x_0.

(iii) f is continuous at x_0 and is of bounded ϕ variation in a neighborhood of x_0 with $\phi \in L^1(0,1)$. Then

$$\lim_{n\to\infty} L_n(w,f,x_0) = f(x_0).$$

4. The results of the previous section could surely be obtained also for Lagrange interpolation based on the zeros of

generalized Hermite polynomials $(w(x) = |x|^{\alpha} \exp(-x^2),\ \alpha > -1,$ $x \in \mathbb{R})$ and Laguerre polynomials $(w(x) = x^{\alpha} \exp(-x),\ \alpha > -1,$ $x \in \mathbb{R}^+)$, respectively. The only new difficulties arise from the fact that these two weights have singularities at zero. Some of the corresponding theorems have been found already, but other problems (like localization) are still open. The next theorem was proved by O. Kis [57]. (For other related results see [71], [77], [79].) Denoting the generalized Hermite weight by w_{α} we have:

THEOREM 17. <u>Let</u> f <u>be uniformly continuous on</u> $\mathbb{R}$ <u>and</u> R <u>be fixed. Then</u>

$$|f(x)-L_n(w_{\alpha},f,x)| = O\left[\log n + |x|^{-\alpha/2}\exp\left(\tfrac{1}{2}x^2\right)\right]\omega_R\left(f,\frac{1}{\sqrt{n}}\right)$$

<u>for</u> $|x| > \dfrac{1}{\sqrt{n}}$ <u>and</u>

$$|f(x)-L_n(w_{\alpha},f,x)| = \begin{cases} O\left[\omega_R\left(f,\frac{1}{\sqrt{n}}\right)\right] & \text{for } -1 < \alpha < 0 \\ O\left[\log n\ \omega_R\left(f,\frac{1}{\sqrt{n}}\right)\right] & \text{for } \alpha = 0 \\ O\left[n^{\alpha/4}\omega_R\left(f,\frac{1}{\sqrt{n}}\right)\right] & \text{for } \alpha > 0 \end{cases}$$

<u>for</u> $|x| \le 1/\sqrt{n}$. <u>These estimates cannot be improved in the sense of Theorems 10 and 11</u>.

5. We conclude Part I by mentioning a few more open problems related to pointwise convergence of Lagrange interpolation.

(a) Let $\operatorname{supp}(w) \subset [-1,1]$, $0 < c_1 \le w(x) \le c_2 < \infty$ for $x \in [-1,1]$. Is it true that

$$p_n(w,1) \le Cp_m(w,1) \quad (n < m)?$$

If $m = n+1$ or $m = n+O(1)$ the answer is yes.

(b) Let $\operatorname{supp}(d\alpha) \subset [-1,1]$. Find conditions, weaker than $(\log \alpha')/\sqrt{1-x^2} \in L^1$, which would assure that

$$\lim_{n\to\infty} \frac{\gamma_{n-1}(d\alpha)}{\gamma_n(d\alpha)} \text{ exists and } = \frac{1}{2}.$$

If $\alpha'(x) > 0$ for almost every $x \in (-1,1)$ and the above limit exists, then it certainly equals $1/2$.

(c) Let w be a "good" weight in some analytic sense with support in $[-1,1]$. Suppose that $w(0) = 0$. How does $L_n(w,f,0)$ behave if f is continuous?

(d) Try to interpolate at the zeros of Pollaczek polynomials. These polynomials are the best known ones with $(\log w)/\sqrt{1-x^2} \notin L^1$ (see [91]).

(e) If $\operatorname{supp}(d\alpha) \subset [-1,1]$ and $\alpha'(x) \geq c > 0$ for almost every $x \in [a,b] \subseteq [-1,1]$ then $|p_n(d\alpha,x)| = o_x(\sqrt{n})$ for $x \in (a,b)$. Will this estimate hold uniformly in $x \in [a_1,b_1] \subset (a,b)$? If $[a,b] = [-1,1]$, α is absolutely continuous and α' is continuous, then the answer is yes.

(f) Let $\operatorname{supp}(d\alpha) \subset [-1,1]$ and ϕ be a continuous, complex valued function. Under what conditions does the limit

$$\lim_{n\to\infty} \frac{\min\limits_{\pi\in\mathbb{P}_n(t)} \left| \int_{-1}^{1} \pi^2(x)\phi(x)d\alpha(x) \right|}{\min\limits_{\pi\in\mathbb{P}_n(t)} \int_{-1}^{1} \pi^2(x)d\alpha(x)}$$

exist and if it exists, what is its value? Here $\mathbb{P}_n(t)$ denotes the set of polynomials π of degree at most n with $\pi(t) = 1$.

(g) What happens if we replace $\mathbb{P}_n(t)$ in (f) by $\mathbb{P}_n^*$, where $\mathbb{P}_n^*$ denotes the class of those polynomials of degree exactly n

whose leading coefficient equals 1. The interesting case is when $\log \alpha' \notin L^1$.

(h) Let $\operatorname{supp}(w) = [-1,1]$. Investigate the expression

$$\frac{1}{p_n(w,1)} L_n(w, fp_{n-1}(w), 1)$$

where f is Riemann integrable.

II. Mean convergence

In this part we will be concerned with mean convergence of Lagrange interpolation. Most of the results mentioned here have been obtained very recently. I hope that in a few years' time the theory of mean convergence of Lagrange interpolation will reach a point where all the results below will be considered trivial. A few words on the notation used: $L^p_{d\alpha} = \{f, \int_{-\infty}^{\infty} |f(t)|^p d\alpha(t) < \infty\}$ for $p > 0$, $\|f\|_{d\alpha,p} = \{\int_{-\infty}^{\infty} |f(t)|^p d\alpha(t)\}^{1/p}$ for $p \geq 1$, $(f,g)_{d\infty} = \int_{-\infty}^{\infty} f(t)g(t)d\alpha(t)$, $\alpha_n(t)$ is defined as a step function, having jumps $\lambda_n(d\alpha, x_{kn}(d\alpha))$ at $x_{kn}(d\alpha)$ and being constant between two consecutive zeros of $p_n(d\alpha)$. Thus, $(f,g)_{d\alpha_n} = \sum_{k=1}^{n} f(x_{kn})g(x_{kn})\lambda_{kn}(d\alpha)$ and $\|f\|_{d\alpha_n,p}$ is defined analogously. The function w_n is defined as $w_n(t) = (\int_{-\infty}^{t} w(x)dx)_n$. $\|L_n(d\alpha)\|_{d\beta,\infty,p}$ denotes the (∞,p) norm of the linear operator $L_n(d\alpha)\colon\ C \to L^p_{d\beta}$ defined by $L_n(d\alpha)f = L_n(d\alpha, f)$.

We shall not bother to compute the (p,p) norm of $L_n(d\alpha)$ for the simple reason that it is never bounded. Let me remark that Part II should be considered as a sequel to R. Askey's paper [3] which helped me understand the importance of this theory.

1. In this section let v denote the Chebyshev weight. Suppose we are asked to prove that the Lagrange interpolation polynomials taken at the Chebyshev abscissas have uniformly bounded $(\infty,2)$ norms as mapping from C to L_v^2. The solution is very simple:

$$\|L_n(v,f)\|_{v,2} = \|f\|_{v_n,2} \leq \|f\|_C \|1\|_{v_n,2} = \|f\|_C \|1\|_{v,2}.$$

The only problem is that this is an essentially L^2 proof and therefore it does not help us to obtain similar inequalities for L_v^p. In a sense it is more natural to compute the L_v^2 norm of $L_n(v,f)$ by using the converse of Hölder's inequality, that is

$$\|L_n(v,f)\|_{v,2} = \sup_{\|g\|_{v,2}\leq 1} (L_n(v,f),g)_v.$$

Denoting by $S_n(v,g)$ the n-th partial sum of the Fourier-Chebyshev series of g, let us remark that $g - S_{n-1}(v,g)$ is orthogonal to $L_n(v,f)$ in L_v^2. Thus, using the Gauss-Jacobi mechanical quadrature formula, we obtain

$$\|L_n(v,f)\|_{v,2} = \sup_{\|g\|_{v,2}\leq 1} (f,S_{n-1}(v,g))_{v_n} \leq$$

$$\leq \|f\|_C \sup_{\|g\|_{v,2}\leq 1} \|S_{n-1}(v,g)\|_{v_n,1} .$$

The next step would be to show that for every polynomial P of degree less than n

$$(2) \quad \|P\|_{v_n,1} \leq C\|P\|_{v,1}$$

with a constant C independent of n and P. Having established such an inequality we could complete our proof by remarking

$$\|S_{n-1}(v,g)\|_{v,1} \leq \|1\|_{v,2}\|g\|_{v,2}.$$

This alternative approach has the advantage of being based on L^1 techniques. L^2 inequalities enter only in the last step to give estimates of the L^1 norms of Fourier-Chebyshev sums (rather than of interpolating polynomials). If (2) holds, then this approach will also yield L^p_v inequalities, since by the celebrated theorem of M. Riesz, the Fourier-Chebyshev sums are uniformly bounded in every L^p_v for $1 < p < \infty$. Let us now return to (2). Following the methods developed in [3], [4], [14], and [99], we represent P as

$$P(x) = (P, K(x,\cdot))_v$$

with a suitably chosen kernel K. Hence

$$\|P\|_{v_n,1} \leq \|P\|_{v,1} \sup_{|x|\leq 1} \|K(x,\cdot)\|_{v_n,1} .$$

Some rather technical arguments show that K may be chosen such that

$$\sup_{|x|\leq 1} \|K(x,\cdot)\|_{v_n,1} \leq C.$$

This proof is fine if we restrict ourselves to considering only the Chebyshev weight or some Jacobi weights (see [4]) but if we try to adapt it for other weights, even the ones which are very close to the Chebyshev weight, we encounter greatly increased technical difficulties which defy all known methods and tricks. Instead we must use a different approach for the proof of (2). (see [62], [63], [68]). By a simple computation

$$\|P\|_{v_n,1} \leq C_1\|P\|_{v,1} + C_2 n^{-1}\|P'\|_1$$

for every polynomial P. In particular, if the degree of P is less than n, then by the Bernstein inequality

$$\|P\|_{v_n,1} \leq (C_1 + C_2)\|P\|_{v,1}.$$

We remark that the second method of proving (2) can also be used in a more general situation.

2. Let us now turn to the formulation of the problem of mean convergence of Lagrange interpolation processes. We fix two weights α and β. Let f be defined on $\Delta_\alpha \cup \operatorname{supp}(d\beta)$ ($\Delta_\alpha = \inf\limits_{\operatorname{supp}(d\alpha) \subset (a,b)} (a,b)$) and let f be finite on Δ_α. If $f \in L^p_{d\beta}$ then also $R_n(d\alpha,f) = f - L_n(d\alpha,f) \in L^p_{d\beta}$ and we can ask whether $R_n(d\alpha,f)$ tends to zero in $L^p_{d\beta}$ as $n \to \infty$.

In this general setting the question seems to be too difficult to be solved. Let us remark that by putting $\beta(x) = \alpha(x) + \delta(x-x_0)$ we find ourselves confronting the problem of pointwise convergence which obviously (?) needs considerations other than just the question of mean convergence. Apart from the case where f is analytic we also cannot expect convergence in $L^p_{d\beta}$ if $\Delta_\beta \not\subset \Delta_\alpha$.

Let A denote the class of those weights α for which

$$\|P\|_{d\alpha_n,1} \leq C\|\|P\|_{d\alpha,1}$$

whenever P is a polynomial of degree less than n. The following theorem is the key in the proof of convergence of Lagrange interpolation in $L^p_{d\beta}$.

THEOREM 1. <u>Let</u> $\alpha \in A$ <u>and let</u> β <u>be absolutely continuous with respect to</u> α. <u>Then for every</u> p <u>such that</u> $1 \leq p < \infty$

$$\|L_n(d\alpha)\|_{d\beta,\infty,p} \leq C\|S_{n-1}(d\alpha)\|_{d\beta,\infty,p}.$$

So far we have not been able to find any weight $\alpha \in A$ with noncompact support. Nevertheless, there exists a rather large class of weights which certainly belongs to A. Let us say that $w \sim w_1$ if $\operatorname{supp}(w) = \operatorname{supp}(w_1)$ and $0 < c_1 \leq w(x)/w_1(x) \leq c_2 < \infty$ for almost every $x \in \operatorname{supp}(w)$.

THEOREM 2. *If* $w \sim w^{(\alpha,\beta)}$ *then* $w \in A$.

This is a consequence of the following more general result.

THEOREM 3. *If* $w \sim w^{(\alpha,\beta)}$ *and* $1 \leq p < \infty$ *then for every polynomial* P *of degree* $m \leq \text{const } n$

$$\|P\|_{d\alpha_n,p} \leq C\|P\|_{d\alpha,p}.$$

Many particular cases of Theorem 3 have been proved in [3], [62], [63], [64], [68]. See [82] for the general case. The proof of Theorem 3 is based on the use of certain Bernstein-Markov inequalities, estimates of the zeros of the corresponding orthonormal polynomials and estimates of Christoffel functions. There are dozens of fine papers treating weighted L^p inequalities for the derivatives of polynomials; see e.g. [15], [16], [51], [61], [65], [84], [86], [88], [90], and [99]. The most satisfactory Bernstein-Markov inequality, however, has been proved by B. Khalilova [53] whose very pretty result goes as follows:

THEOREM 4. *If* $P \in \mathbb{P}_n$ *then*

$$\|P'\|_{w^{(\alpha+p/2,\beta+p/2)},p} \leq Cn\|P\|_{w^{(\alpha,\beta)},p}$$

for $\alpha,\beta > -1$ *and* $p \geq 1$.

The next theorem is needed to deal with the Christoffel functions and the zeros of orthogonal polynomials in the proof of Theorem 3. (See [21], [24], [32], [76], [82], [91].)

THEOREM 5. *Let* $w \sim w^{(\alpha,\beta)}$. *Then*

$$\lambda_n(w,x) \sim \frac{1}{n}\left[\sqrt{1-x} + \frac{1}{n}\right]^{2\alpha+1}\left[\sqrt{1+x} + \frac{1}{n}\right]^{2\beta+1}$$

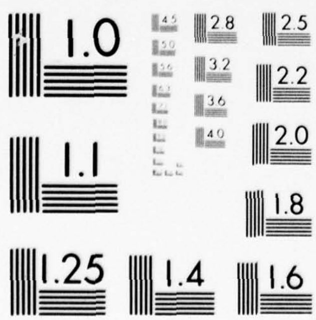
for $|x| \leq 1$ *and*

$$\theta_{k+1,n}(w) - \theta_{kn}(w) \sim \frac{1}{n}$$

AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1
APPROXIMATION THEORY II
NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.
UNCLASSIFIED AFOSR TR-76-1242 N/L

3 OF 7
AD
A033032

033032



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

<u>for</u> $k = 0,1,\ldots,n$ <u>where</u> $x_{kn}(w) = \cos\theta_{kn}(w)$ <u>with</u> $x_{0n}(w) = 1$ <u>and</u> $x_{n+1,n}(w) = -1$.

Having Theorems 1 and 2, the next step is to apply results on weighted mean convergence of orthogonal Fourier series. Such questions will be discussed in the next section.

3. If $\operatorname{supp}(d\alpha)$ is compact then for every $f \in L^2_{d\alpha}$

$$S_n(d\alpha,f) = \sum_{k=0}^{n} (f,p_k(d\alpha))_{d\alpha}p_k(d\alpha)$$

converges to f in $L^2_{d\alpha}$ when $n \to \infty$. Much less can be said about the convergence of $S_n(d\alpha,f)$ in $L^p_{d\beta}$. In the case where α and β are Jacobi weights (not necessarily the same) or are close to Jacobi weights in an analytic sense, the problem of convergence of $S_n(d\alpha,f)$ in $L^p_{d\beta}$ has now been almost completely solved. On the other hand, when, for instance, α is absolutely continuous and α' is the Chebyshev weight multiplied by a function which is merely continuous, nothing is known about the convergence of $S_n(d\alpha,f)$ in $L^p_{d\alpha}$. For known results, we refer the reader to [2], [6], [7], [8], [9], [10], [48], [50], [66], [67], [83], [85], [97], and [99].

For our purposes we need theorems of this type which can be combined with Theorems 1 and 2 to give mean convergence of Lagrange interpolation. The strongest available result is that of V. M. Badkov [11]. He has dealt with the case when $\operatorname{supp}(w) = [-1,1]$,

$$w(t) = h(t)w^{(\alpha,\beta)}(t)\prod_{j=1}^{m}|t-x_j|^{\gamma_j},$$

$$h(t) > 0,\ h(t) \in C,\ \omega(h,\delta)\delta^{-1} \in L^1,$$

$$\alpha,\beta,\gamma_j > -1 \quad (j = 1,2,\ldots,m), \quad -1 < x_1 < \ldots < x_m < 1.$$

For such weights w he considered the mean convergence of

$S_n(w,f)$ in L^p_v where

$$v(t) = w^{(A,B)}(t) \prod_{j=1}^{m} |t-x_j|^{\Gamma_j},$$

A, B, $\Gamma_j > -1$ $(j = 1,2,\dots,m)$, x_j are the same as above.

THEOREM 6. Let $1 < p < \infty$. Then $\|S_n(w)\|_{v,p,p} \leq C$ for every n iff

$$(3) \quad \sup_{n\geq 0}\left\{\|p_n(w)\|_{v,p}\right\} < \infty, \quad \sup_{n\geq 0}\left\{\|p_n(w)w\|_{v^{1/1-p},p/p-1}\right\} < \infty.$$

THEOREM 7. Conditions (3) are equivalent to the system of inequalities.

$$\left|\frac{A+1}{p} - \frac{\alpha+1}{2}\right| < \min\left\{\frac{1}{4}, \frac{\alpha+1}{2}\right\},$$

$$\left|\frac{B+1}{p} - \frac{\beta+1}{2}\right| < \min\left\{\frac{1}{4}, \frac{\beta+1}{2}\right\},$$

$$\left|\frac{\Gamma_j+1}{p} - \frac{\gamma_j+1}{2}\right| < \min\left\{\frac{1}{2}, \frac{\gamma_j+1}{2}\right\}, \qquad j = 1,2,\dots,m.$$

4. From this point it becomes easy to obtain results on mean convergence of Lagrange interpolation from the results mentioned above. Let us write $w \approx w^{(\alpha,\beta)}$ if $\mathrm{supp}(w) = [-1,1]$, $w \sim w^{(\alpha,\beta)}$, $w/w^{(\alpha/\beta)} \in C$, $w/w^{(\alpha/\beta)} > 0$ and the modulus of continuity $\omega(\delta)$ of $w/w^{(\alpha,\beta)}$ satisfies the condition

$$\int_0^1 \frac{\omega(\delta)}{\delta}\, d\delta < \infty.$$

THEOREM 8. Let $w \approx w^{(\alpha,\beta)}$ and $v = uw^{(a,b)}$ where u^{ϵ} is integrable for some $\epsilon > 1$. For every $f \in C[-1,1]$ $L_n(w,f)$ converges to f in $L^p_v[-1,1]$ when $n \to \infty$ if

(i) $\max(\alpha,\beta) \leq -\frac{1}{2}$, $a = 0$, $b = 0$ and $p > 0$

(ii) $\min(\alpha,\beta) > -\frac{1}{2}$, u is bounded in some neighborhoods

of -1 *and* 1 *and*

$$0 < p < \min\left\{\frac{4(a+1)}{2\alpha+1}, \frac{4(b+1)}{2\beta+1}\right\}.$$

(iii) $\alpha \leq -\frac{1}{2} < \beta$, u *is bounded in a neighborhood of* -1, $a = 0$, $b > \frac{2\beta-3}{4}$ *and*

$$0 < p < \frac{4(b+1)}{2\beta+1}.$$

(iv) $\beta \leq -\frac{1}{2} < \alpha$, u *is bounded in a neighborhood of* 1, $b = 0$ *and*

$$0 < p < \frac{4(a+1)}{2\alpha+1}.$$

Note that the conditions imposed on p are also necessary. This is clear from the following theorem.

THEOREM 9. *Let* $w \approx w^{(\alpha,\beta)}$ *and* $v = uw^{(a,b)}$. *Suppose that* $L_n(w,f)$ *tends to* f *in* L_v^p *for every* $f \in C$. *If* $\alpha > -\frac{1}{2}$ *and* u^{-1} *is bounded in a neighborhood of* 1, *then necessarily*

$$p < \frac{4(a+1)}{2\alpha+1}$$

and similarly if $\beta > -\frac{1}{2}$ *and* u^{-1} *is bounded in a neighborhood of* -1 *then*

$$p < \frac{4(b+1)}{2\beta+1}.$$

The method which I used to prove Theorem 8 cannot be applied to prove convergence of $L_n(d\alpha,f)$ to f in $L^p_{d\beta}$ when $f \in L^p_{d\beta}$ in the Riemann-Stieltjes sense. However, if $\alpha \equiv \beta$, this method will probably work. Theorems 8 and 9 have been proved in [82]. They contain earlier results proved in [3], [4], [19], [22], [23], [52], [63], [92], and also the classical L^2 theorems of P. Erdös-P. Turán [21] and J. Shohat [89].

It should be possible to show that the weight

$$w(t) = h(t)w^{(\alpha,\beta)}(t) \prod_{j=1}^{m} |t-x_j|^{\gamma_j}$$

$$(\gamma_j > -1,\ x_j \in (-1,1),\ 0 < c_1 \le h(t) \le c_2 < \infty)$$

belongs to A and thus to extend Theorem 8 for such weights. A further step would be to investigate the convergence of $L_n(d\alpha, f)$ in $L^p_{\alpha\beta}$ when $f \in L^p_{\alpha\beta}$ in the improper Riemann-Stieltjes sense. G. Freud's book [32] is a good starting point for such studies. For those who prefer the method of "nice kernel functions" to that of Bernstein-Markov inequalities I recommend the recent monograph [5] of R. Askey.

As mentioned above, all the weights we know to be in A have compact support. For this reason the above machinery does not work when supp($d\alpha$) is not compact. Nevertheless, after a little modification it becomes applicable for certain weights with noncompact support. This question will be discussed in the following section.

5. Recent investigations of mean convergence of Lagrange interpolation processes corresponding to weights with noncompact support arose from the physical problem of obtaining approximate expressions for the Fourier transforms of functions which are only known experimentally. One reasonable approach to this question is to replace the function, which is supposedly "known" only at certain point, by an interpolating polynomial, then to take the Fourier transform of this polynomial and to see whether it is in fact a good approximation to the Fourier transform of the function itself. As was shown in [13], if the function is given at the Laguerre nodes $(w(x) = \exp(-x),\ x \in \mathbb{R}^+)$ this method works well. To be more precise,

$$\lim_{n\to\infty} \int_0^\infty L_n(w,f,t)e^{-t+ixt}dt = \int_0^\infty f(t)e^{-t+ixt}dt$$

uniformly in $x \in \mathbb{R}$ provided that f is continuous and $\lim_{t\to+\infty} f(t)\exp(-\epsilon t) = 0$ for some $\epsilon < \frac{1}{2}$. This result gave rise to [12] and [26] where the problem of convergence in L^2 was thoroughly investigated.

J. Balázs and P. Turán [12] dealt with the case where α is absolutely continuous and $w = \alpha'$ can be represented in the form $w = h/g$ where $h \in L^1(\mathbb{R})$ and g is an even, infinitely many times differentiable function which satisfies the conditions: $g^{(k)}(x) > 0$ $(k = 0,2,\ldots)$, $\log g$ is convex function of $\log x$ and

$$\int_0^\infty \frac{\log g(t)}{1+t^2} = +\infty.$$

THEOREM 10. <u>If</u> $f \in C(\mathbb{R})$ <u>and</u> $\lim_{x\to\infty} f(x)/\sqrt{g(x)} = 0$ <u>then</u> $L_n(d\alpha, f)$ <u>converges to</u> f <u>in</u> $L^2_{d\alpha}(\mathbb{R})$ <u>when</u> $n \to \infty$.

The following special case of Theorem 10 is of particular interest.

THEOREM 11. <u>Let</u> $f \in C(\mathbb{R})$ <u>and</u> $f(x) = O[\exp(\epsilon x^2)]$ <u>when</u> $x \to \infty$ <u>with some</u> $\epsilon < \frac{1}{2}$. <u>Then the corresponding Lagrange interpolation polynomials taken at the Hermite abscissas converge to</u> f <u>in</u> $L^2_w(\mathbb{R})$ <u>with</u> $w(x) = \exp(-x^2)$.

In order to formulate G. Freud's result in [26] we have to introduce a definition. We say that the weight α belongs to U if the equalities

$$\int_{-\infty}^{\infty} x^n d\alpha(x) = \int_{-\infty}^{\infty} x^n d\beta(x) \qquad (n = 0,1,\ldots)$$

imply that $\alpha(x) - \alpha(-\infty) = \beta(x) - \beta(-\infty)$ at every point x where α or β is continuous.

THEOREM 12. <u>Let</u> $\alpha \in U$ <u>and let</u> G <u>be a function which is</u>

<u>infinitely many times differentiable and satisfies the conditions</u> $G^{(k)}(x) \geq 0 \quad (k = 0,2,\dots)$ <u>and</u> $G \in L^1_{d\alpha}$. <u>Then for every function</u> f <u>with</u> $\lim_{x\to\infty} f(x)/\sqrt{G(x)} = 0$, $L_n(d\alpha, f)$ <u>tends to</u> f <u>in</u> $L^2_{d\alpha}(\mathbb{R})$ <u>provided that</u> $f \in L^2_{d\alpha}(\mathbb{R})$.

Here, and in the sequel $L^p_{d\alpha}(\mathbb{R})$ should be understood in the improper Riemann-Stieltjes sense; that is, $f \in L^p_{d\alpha}(\mathbb{R})$ if

$$(R\text{-}S) \int_{-a}^{a} |f(t)|^p d\alpha(t)$$

exists for every $a > 0$ and

$$\lim_{a\to\infty} (R\text{-}S) \int_{-a}^{a} |f(t)|^p d\alpha(t) < \infty.$$

We shall not go into the details of the proofs of the results in this section; we only mention that they are essentially L^2 proofs and cannot be generalized to handle the case of convergence in L^p. (See also [93].) Both P. Turán [92] and G. Freud [32] have raised the question of what happens in L^p.

Of course, there is virtually no hope of obtaining L^p convergence theorems for general weights while the classical cases, namely Hermite and Laguerre weights, remain uninvestigated. R. Askey [3], recalling an old idea of J. Marcinkiewicz [63], realized that the question of mean convergence of Lagrange interpolation can be reduced to investigation of mean convergence of orthogonal Fourier series (see Theorem 1). It turns out to be unfortunate that R. Askey knows a very great deal about mean convergence of orthogonal Fourier series and he knew well that there exist no good convergence theorems in L^p for Fourier-Hermite and Fourier-Laguerre series, respectively. For this reason he wrote in [3] that "the lack of nice theorems (for L^p convergence) suggests that there are only fairly weak results to be obtained for Lagrange interpolation at the zeros of the

Laguerre or Hermite polynomials. Turán raised this question... and I too would like to see some results on this question. However, I am afraid they will be weaker than one might have expected!"

Askey was right only partially. There are both weak and strong theorems for mean convergence of Lagrange interpolation taken at the Hermite or Laguerre abscissas.

6. Throughout this section w will denote the Hermite weight; that is, $w(x) = \exp(-x^2)$, and let $w_{(p)} = w^{p/2}$. The weak result referred to above is the following:

THEOREM 13. _Let_ $p > q \geq 1$. _Then there exists a continuous function_ f _supported in_ $[-1,1]$ _such that_

$$\limsup_{n \to \infty} \|L_n(w,f)\|_{w_{(q)},p} = \infty,$$

in particular, for every $p > 2$ _one can find a continuous_ f _with support in_ $[-1,1]$ _such that_

$$\limsup \|L_n(w,f)\|_{w,p} = \infty.$$

The proof of this theorem is surprisingly easy. Since we can consider $L_n(w)$ as linear operators from $C[-1/2,1/2]$ into $L^p_{w_{(q)}}(\mathbb{R})$, by the Banach-Steinhaus theorem it is enough to construct a sequence $\{f_n\}$ with $f_n \in C[-1/2,1/2]$, $\|f_n\|_c \leq 1$ and

$$\limsup_{n \to \infty} \|L_n(w,f_n)\|_{w_{(q)},p} = \infty.$$

The existence of the above sequence $\{f_n\}$ follows immediately from the next two theorems.

THEOREM 14. _For every_ n _there exists a function_ $f_n \in C[-1/2,1/2]$ _with_ $\|f_n\|_c = 1$ _and a number_ $x_n \in (\sqrt{n} - 3, \sqrt{n} + 3)$ _such that_

$$\sqrt{n}\, w_{(1)}(x_n)\,|L_n(w,f_n,x_n)| \geq C.$$

THEOREM 15. <u>Let</u> $P \in \mathbb{P}_n$. <u>Then for every</u> $p \geq 1$

$$\|Pw_{(2/p)}\|_C \leq Cn^{1/2p}\|P\|_{w,p}.$$

Theorem 14 has been proved in [29]. Theorem 15 is obvious when p is an even natural integer and for the other values of p it can be proved by an interpolation (not Lagrange but Riesz-Thorin) argument. Theorem 13 suggests that if we want to have convergence in L^p we should consider convergence in $L^p_{w_{(p)}}(\mathbb{R})$ rather than in $L^p_w(\mathbb{R})$.

THEOREM 16. <u>Let</u> $f \in C(\mathbb{R})$ <u>and satisfy the condition</u>

(4) $\lim_{x\to\infty} f(x)xw_{(1)}(x) = 0.$

<u>Then for every</u> $p > 1$,

(5) $\lim_{n\to\infty} \|L_n(w,f) - f\|_{w_{(p)},p} = 0$.

It is quite possible that a condition weaker than (4) could also imply (5). On the other hand, it should be possible to show that the conditions

$$f \in L^p_{w_{(p)}}(\mathbb{R}) \cap C(\mathbb{R}),\ \lim_{x\to\infty} f(x)w_{(1)}(x) = 0$$

are too weak to imply (5). The proof of Theorem 16 is rather complicated; it is based on the following two theorems.

THEOREM 17. <u>Let</u> $P \in \mathbb{P}_n$ <u>and</u> $p \geq 1$. <u>Then</u>

$$\|P\|_{w_{(p)},p} \leq C_1\|1^*_{(n)}P\|_{w_{(p)},p},$$

<u>where</u> $1^*_{(n)}$ <u>is the characteristic function of</u> $[-c_2\sqrt{n},\ c_2\sqrt{n}\,]$ <u>with an appropriate constant</u> c_2 <u>and</u>

$$\|P'\|_{w_{(p)},p} \leq C\sqrt{n}\,\|P\|_{w_{(p)},p}.$$

THEOREM 18. Let $P \in \mathbb{P}_m$ with $m \leq \text{const}\, n$. Then for every $p \geq 1$, $a \in \mathbb{R}$, $0 \leq b < p^{-1}$

$$(6) \quad \|PG_{a,b}\|_{w_n,p} \leq C\|PG_{a+2/p,b}\|_{w,p}$$

and

$$(7) \quad \|PG_{a,b}1_{(n)}\|_{w_n,p} \leq C\|PG_{a,b}\|_{w,p}$$

where $G_{a,b}(x) = (1+|x|)^a \exp(bx^2)$ and $1_{(n)}$ is the characteristic function of $[-\sqrt{n}, \sqrt{n}]$.

The very important Theorem 17 has been proved by G. Freud [35]. It has nice applications in the theory of weighted L^p approximation which has been developed by G. Freud in [33]-[41]. Let us remark that if we put $p = 2$, $a = b = 0$ in (6) we get an inequality which is not sharp for polynomials of degree less than n, since on the right side we have an extra factor $(1+|x|)^2$. Hence it is possible that this factor can also be omitted in the general case, that is, $w \in A$. Instead of proving Theorem 16 let us discuss an essential step in its proof which we formulate as

THEOREM 19. Let

$$M = \sup_{x\in\mathbb{R}} (1 + |x|)w_{(1)}(x)|f(x)| < \infty.$$

Then for every $p > 1$

$$\|L_n(w,1_{(n)}f)1_{(n)}\|_{w_{(p)},p} \leq CM.$$

Proof. We have

$$I = \|L_n(w,1_{(n)}f)1_{(n)}\|_{w_{(p)},p} = \sup_{\|g\|_{w,q}\leq 1} (L_n(w,1_{(n)},f),g^*)_w$$

where $q = p/(p-1)$ and $g^* = 1_{(n)}gw_{(p-2/p)}$. Hence

$$I = \sup_{\|g\|_{w,q} \leq 1} (L_n(w, 1_{(n)} f),\ S_n(w, g^*))_w$$

$$= \sup_{\|g\|_{w,q} \leq 1} (L_n(w, 1_{(n)} f),\ S_n(w, g^*))_{w_n}$$

$$= M \sup_{\|g\|_{w,q} \leq 1} \|1_{(n)} G_{-1,1/2} S_n(w, g^*)\|_{w_n, 1}$$

and by (7)

$$I \leq CM \sup_{\|g\|_{w,q} \leq 1} \|G_{-1,1/2} S_n(w, g^*)\|_{w,1}.$$

From this

$$I \leq CM \sup_{\|g\|_{w,q} \leq 1} \sup_{\|h\|_c \leq 1} (G_{-1,1/2} S_n(w, g^*), h)_w$$

$$= \sup_{\|g\|_{w,q} \leq 1} \sup_{\|h\|_c \leq 1} (g^*, S_n(w, G_{-1,1/2} h))_w.$$

By Hölder's inequality

$$I \leq CM \sup_{\|h\|_c \leq 1} \|1_{(n)} S_n(W, G_{-1,1/2} h)\|_{w_{(p)}, p}.$$

Here the right-hand side can easily be estimated remembering that for every $\emptyset \in L^1_w(\mathbb{R})$

$$|S_n(w, \emptyset)| = \frac{\gamma_n(w)}{\gamma_{n+1}(w)} \left| p_n(w) H(\emptyset p_{n+1}(w) w) - p_{n+1}(w) H(\emptyset p_n(w) w) \right|$$

where H denotes the Hilbert transform. To finish the proof we use some well-known properties of Hermite polynomials and obtain $I \leq CM$.

We refer the reader interested in mean convergence of Fourier-Hermite series to [7], [67], [83], and [85].

7. Similar theorems can be proved for the generalized Hermite or Laguerre weights. These results will soon be

published. However, attempts to extend these results to weights other than the classical ones come up against serious difficulties. These difficulties arise because not much is known about the behavior of orthogonal polynomials corresponding to weights with noncompact support. For certain weights, like $|x|^\alpha \exp(-|x|^\beta)$ and $\exp(g(x))$ (where g is a "good" function), some properties of the corresponding orthogonal polynomials are already known. These properties include exact estimates of the Christoffel functions, of distances between consecutive zeros and of leading coefficients (see [18], [20], [42]-[47], [75], and [76]). A next step would be to find asymptotic formulae or at least useful bounds for the orthogonal polynomials. In particular, it would be interesting to have a proof of the following conjecture.

CONJECTURE. Let $w(x) = \exp(-|x|^\alpha)$ $(\alpha > 0,\ x \in \mathbb{R})$. Then

$$p_n^2(w,x)w(x) \le C \qquad (x \in \mathbb{R})$$

and there exists a constant c_1 such that

$$p_n^2(w,x)w(x) \le Cn^{-1/\alpha} \qquad (|x| \le c_1 n^{1/\alpha}),$$

or, more generally, if $w(x) = \exp(-|x|^\alpha)h(x)$ where h is a "very good" strictly positive and bounded function, then the above inequalities hold.

Even for just the special case $\alpha = 2$, a proof of the second form of the conjecture would be of considerable interest.

8. We conclude this survey with a few open problems.

(a) Find conditions assuring the existence of a $p > 2$ such that $p_n(w)$ is uniformly bounded in L_w^p.

(b) Investigate mean convergence of $L_n^{(k)}(d\alpha, f)$ to $f^{(k)}$, provided that $f^{(k)} \in C$.

(c) Determine the asymptotic value of $\sup_{f\in M} \|L_n(d\alpha,f)-f\|_{d\beta,p}$

for $n \to \infty$ if M is a compact subset of C.

(d) Estimate $x_{k+1,n}(w) - x_{kn}(w)$ when w is between two different Jacobi weights.

(e) Investigate the mean convergence of $L_n^*(d\alpha, f)$ which coincides with f at $\{x_{kn}(d\alpha)\} \cup \{\pm 1\}$.

(f) Does A (see Theorem 1) contain any weight supported on two disjoint intervals?

(g) Let $\mathrm{supp}(w) = [-1,1]$. If $(\log w)/\sqrt{1-x^2} \in L^1$ then for every linear functional F on L^1

$$\lim_{n\to\infty} F(p_n^2(w)w) = F(v)$$

where v is the Chebyshev weight. Is the condition $\log w \in L^1$ necessary?

(h) Give exact estimates for $\|L_n(w^{(\alpha,\beta)}\|_{w^{(a,b)},\infty,p}$ when it is not bounded.

(i) If $\mathrm{supp}(d\alpha) \subset [-1,1]$ and $(\log \alpha')/\sqrt{1-x^2} \in L^1$ then $\lambda_n(d\alpha,x)p_n^2(d\alpha,x) = \sigma_x(1)$ for $x \in (-1,1)$. When will this extimate hold uniformly on $[a,b] \subset (-1,1)$?

Finally, I would like to thank Richard Askey and Michael Civikel for having kindly consented to read a draft version of this survey and to offer several helpful suggestions for improving the presentation.

References

1. Agahanov, S.A., Estimation of the Lebesgue function of interpolation taken at Jacobi-abscissas, Izv. Vyssh. Uchebn. Zaved. Matematika 11 (1967), 3-6. (In Russian.)

2. Askey, R., Norm inequalities for some orthogonal series, Bull. Amer. Math. Soc. 73 (1966), 802-823.

3. Askey, R., Mean convergence of orthogonal series and Lagrange interpolation, Acta Math. Acad. Sci. Hungar. 23

(1972), 71-85.

4. Askey, R., Summability of Jacobi series, Trans. Amer. Math. Soc. 179 (1973), 71-84.

5. Askey, R., Orthogonal Polynomials and Special Functions, Soc. for Ind. and Applied Math., Philadelphia, Penn., 1975.

6. Askey, R. and I.I, Hirschman Jr., Mean summability for ultraspherical polynomials, Math. Scand. 12 (1963), 167-177.

7. Askey, R. and S. Wainger, Mean convergence of expansions in Laguerre and Hermite series, Amer. J. Math. 87 (1965), 695-708.

8. Badkov, V.M., Equiconvergence of Fourier series in orthogonal polynomials, Mat. Zametki 5 (1969), 285-295. (In Russian.)

9. Badkov, V.M., Mean convergence of orthogonal Fourier series, Mat. Zametki 14 (1973), 161-172. (In Russian.)

10. Badkov, V.M., Boundness in mean of orthogonal polynomials, Mat. Zametki 13 (1973), 759-770. (In Russian.)

11. Badkov, V.M., Mean convergence and almost everywhere convergence of orthogonal Fourier series, Mat. Sbornik 95(137) (1974), 229-262. (In Russian.)

12. Balázs, J. and P. Turán, Notes on interpolation, VIII, Acta Math Acad. Sci. Hungar. 12 (1961), 469-474.

13. Balázs, J. and P. Turán, Notes on interpolation, IX, Acta Math Acad. Sci. Hungar. 16 (1965), 215-220.

14. Berman L.D., Some inequalities and their application to interpolation theory, Investigations on Modern Problems of Constr. Func. Th., edited by V.I.Smirnov, Moscow, 1961, 116-122. (In Russian.)

15. Daugavet, I.K. and S.Z. Rafalson, Markov-Nikol'skii type estimates for algebraic polynomials, Vestnik Leningrad. Univ. Mat. Meh. Astronom. 11 (1972), 13-25. (In Russian.)

16. Dzjadyk, V.K., On constructive characterization of functions satisfying the Lipschitz condition on a finite interval of the real line, Izv. Akad. Nauk SSSR, Ser. Mat. 20 (1956), 623-643. (In Russian.)

17. Erdős, P., Problems and results on the convergence and divergence properties of the Lagrange interpolation polynomials and some extremal problems, Mathematica (Cluj) (33)19 (1968), 65-73.

18. Erdős, P., On the distribution of the roots of orthogonal polynomials, Proc. of the Conf. on Constr. Th. of Func., edited by G. Alexits, Budapest, 1972, 145-150.

19. Erdös, P. and E. Feldheim, Sur le mode de convergence pour l'interpolation de Lagrange, Comptes Rendus Paris 203 (1936), 913-915.

20. Erdös, P. and G. Freud, On the distribution function of the zeros of orthogonal polynomials, Proc. London. Math. Soc., 29 (1974), 521-537.

21. Erdös, P. and P. Turán, On interpolation, I,II,III, Annals of Math. 38 (1937), 142-155, 39 (1938), 703-724, 41 (1940), 510-553.

22. Feldheim, E., Sur le mode de convergence dans l'interpolation de Lagrange, Doklady AN SSSR 14 (1937), 327-331.

23. Feldheim, E., Théorie de la convergence des procédes d' interpolation et de quadrature mécanique, Mémorial des Sc. Math. 95 (1939).

24. Freud, G., Über einen Satz von P. Erdös and P. Turán, Acta Math. Acad. Sci. Hungar. 4 (1953) 255-266.

25. Freud, G., Über differenzierte Folgen der Lagrangeschen Interpolation, Acta Math. Acad. Sci. Hungar. 6 (1955) 467-473.

26. Freud, G., Über die Konvergenz in Mittel von Lagrangeschen Interpolationspolynomfolgen, Acta Math. Acad. Sci. Hungar. 13 (1962), 257-268.

27. Freud, G., Convergence of Lagrange interpolation processes on infinite interval, Mat. Lapok 18 (1967), 289-292. (In Hungarian.)

28. Freud, G., Über die Lokalisationseigenschaften und die starke (C,1)-Summation Lagrangescher Interpolationsfolgen, Publ. Math. 15 (1968), 293-301.

29. Freud, G., Lagrangesche Interpolation über die Nullstellen der Hermiteschen Orthogonalpolynome, Studia Sci. Math. Hungar. 4 (1969), 179-190.

30. Freud, G., Über eine Klasse Lagrangeschen Interpolationsverfahren, Studia Sci. Math. Hungar. 3 (1968), 249-255.

31. Freud, G., Ein Beitrag zur Theorie des Lagrangeschen Interpolationverfahrens, Studia Sci. Math. Hungar. 4 (1969), 379-384.

32. Freud, G., Orthogonal Polynomials, Pergamon Press, New York, 1971.

33. Freud, G., On weighted polynomial approximation on the whole real axis, Acta Math. Acad. Sci. Hungar. 20 (1969) 223-225.

34. Freud, G., On weighted approximation by polynomials on the real axis, Doklady AN SSSR 191 (1970), 293-294. (In Russian.)

35. Freud, G., On a Markov-type inequality, Doklady AN SSSR 197 (1971), 790-793. (In Russian.)

36. Freud, G., A contribution to the problem of weighted polynomial approximation, Linear Operators and Approximation, edited by P.L. Butzer, J.-P. Kahane, B. Sz.-Nagy, Birkhäuser Verlag, 1971, 431-447.

37. Freud, G., On two polynomial inequalities, I,II, Acta Math. Acad. Sci. Hungar. 22 (1971), 109-116, 23 (1972), 137-145.

38. Freud, G., On approximation by polynomials with weight $\exp(-\frac{1}{2}x^2)$, Doklady AN SSSR 201 (1971), 1292-1294. (In Russian.)

39. Freud, G., On direct and converse theorems in the theory of weighted polynomial approximation, Math. Z. 126 (1972), 123-134.

40. Freud, G., On weighted L_1-approximation by polynomials, Studia Math. 46 (1973), 125-133.

41. Freud, G., On the converse theorems of weighted polynomial approximation, Acta Math. Acad. Sci. Hungar. 24 (1973), 389-397.

42. Freud, G., On the greatest zero of an orthogonal polynomial, I,II, Acta Sci. Math. Szeged, 34 (1973), 91-97, 35 (1974), 49-54.

43. Freud, G., An estimation of the greatest zeros of orthogonal polynomials, Acta Math. Acad. Sci. Hungar. 25 (1974), 99-107.

44. Freud, G., On the zeros of orthogonal polynomials with respect to measures with noncompact support, Proc. of the Coll. on Constr. Func. Theory, Cluj, 1973.

45. Freud, G., On the coefficient in the recursion formula of orthogonal polynomials, Studia Sci. Math. Hungar. (to appear).

46. Freud, G. and G.P. Nevai, Weighted L_1 and one-sided weighted L_1 polynomial approximation on the real line, Magyar Tudományos Akadémia III. Oszt. Kōzl. 21 (1973), 485-502.

47. Freud, G. and G.P. Nevai, Über einseitige Approximation durch Polynome, III, Acta Sci. Math. Szeged 35 (1973), 65-72.

48. Generozov, V.L., On L_p convergence of expansions in eigenfunctions of a Stourme-Liouville problem, Mat. Zametki 3

(1968), 683-691. (In Russian.)

49. Geronimus, Ja. L., Convergence of Lagrange interpolation taken at zeros of orthogonal polynomials, Izv. Akad. Nauk SSSR, Ser. Mat. 27 (1963), 529-560. (In Russian.)

50. Gilbert, J.E., Maximal theorems for some orthogonal series, I,II, Trans. Amer. Math. Soc. 145 (1969), 495-516 and J. Math. Anal. Appl. 31 (1970), 349-368.

51. Hille, E., G. Szegö and J.D. Tamarkin, On some generalization of a theorem of A. Markoff, Duke Math. J. 3 (1937), 729-739.

52. Holló, A., On mechanical quadrature, Thesis, Budapest, 1939.

53. Khalilova, V.A., On some estimates for polynomials, Izv. AN Azerbaĭdžan SSR. Ser. Fiz-Teh. Mat. 2 (1974), 46-55.

54. Kis, O., Notes on the speed of convergence of Lagrange interpolation, Annales Univ. Sci. Budapest, Sec. Math. 11 (1968) 27-40. (In Russian.)

55. Kis, O., Notes on estimation of deviation of interpolation, Studia Sci. Math. Hungar. 4 (1969), 279-289. (In Russian.)

56. Kis, O., Notes on estimation of deviation of trigonometric and Lagrange interpolation, Proc. of the Conf. on Constr. Th. of Func., edited by G. Alexits, Budapest, 1972, 255-259. (In Russian.)

57. Kis, O., Lagrange interpolation at the zeros of Sonine-Markov polynomials, Acta Math. Acad. Sci. Hungar. 23 (1972) 389-417. (In Russian.)

58. Kis, O. and J. Szabados, Notes on the convergence of Lagrange interpolation, Acta Math. Acad. Sci. Hungar. 16 (1965), 389-430. (In Russian.)

59. Krylov, V.I., Convergence of algebraic interpolation in classes of differentiable functions, Doklady AN SSSR 105 (1955), 214-217. (In Russian.)

60. Krylov, V.I., Convergence of algebraic interpolation taken at the Tschebisev abscissas for absolutely continuous functions and functions of bounded variation, Doklady AN SSSR 107 (1956), 362-365. (In Russian.)

61. Lebed', G.K., Inequalities for polynomials and their derivatives, Doklady AN SSSR 117 (1957), 570-572. (In Russian.)

62. Lozinsky, S.M., Convergence and summability of Fourier series and interpolation processes, Mat. Sbornik 14(56) (1944), 175-268.

63. Marcinkiewicz, J., Sur l'interpolation, I,II, Studia Math. 6 (1936), 1-17, 67-81.

64. Marcinkiewicz, J. and A. Zygmund, Mean values of trigonometric polynomials, Fund. Math. 28 (1937), 131-166.

65. Michlin, S.G., On Markov-type inequalities, Vestnik Leningrad. Univ. Mat. Meh. Astronom. 13 (1969), 54-68.

66. Muckenhoupt, B., Mean convergence of Jacobi series, Proc. Amer. Math. Soc. 23 (1969), 306-310.

67. Muckenhoupt, B., Mean convergence of Hermite and Laguerre series, I,II, Trans. Amer. Math. Soc. 147 (1970), 419-431 and 433-460.

68. Natanson, G.I., On a theorem of S.M. Lozinsky, Doklady AN SSSR 117 (1957), 32-35. (In Russian.)

69. Natanson, G.I., Generalization of a theorem of S.M. Lozinsky for nontriangular methods of summation, Dolady AN SSSR 218 (1961), 141-155. (In Russian.)

70. Natanson, G.I., Two-sided estimation of the Lebesgue function of Lagrange interpolation process taken at Jacobi abscissas, Izv. Vyssh. Ucebn. Zaved. Matematika 11 (1967) 67-74. (In Russian.)

71. Nevai, G.P., Lagrange interpolation at the zeros of Laguerre polynomials, Mat. Lapok 22 (1971), 149-164. (In Hungarian.)

72. Nevai, G.P., Lagrange interpolation on the real line, Mat. Lapok 23 (1972), 143-146. (In Hungarian.)

73. Nevai, G.P., On an inequality of G. Freud, Annales Univ. Sci. Budapest, sec. math. 16 (1973), 87-92. (In Russian.)

74. Nevai, G.P., On Lagrange interpolation taken at the zeros of Hermite polynomials, Acta Math. Acad. Sci. Hungar. 24 (1973), 209-213. (In Russian.)

75. Nevai, G.P., Some properties of orthonormal polynomials with respect to weight $(1+x^{2k})^{\alpha}\exp(-x^{2k})$ and their application in approximation theory, Doklady AN SSSR 211 (1973), 784-786. (In Russian.)

76. Nevai, G.P., Polynomials which are orthonormal on the real axis with respect to weight $|x|^{\alpha}\exp(-|x|^{\beta})$, I, Acta Math. Acad. Sci. Hungar. 24 (1973), 335-342. (In Russian.)

77. Nevai, G.P., Convergence of Lagrange interpolation based on the Laguerre nodes, Publ. Math. Debrecen 20 (1973), 235-239. (In Russian.)

78. Nevai, G.P., On the Dini-Lipschitz convergence test, Acta

Math. Acad. Sci. Hungar. 24 (1973), 349-351. (In Russian.)

79. Nevai, G.P., Notes on interpolation, Acta Math. Acad. Sci. Hungar. 25 (1974), 123-144. (In Russian.)

80. Nevai, G.P., Localization theorems on convergence of Lagrange interpolation taken at the Hermite abscissas, Acta Math. Acad. Sci. Hungar. 25 (1974), 341-361. (In Russian.)

81. Nevai, G.P., Fourier series and interpolation, I,II, Mat. Lapok 23 (1972), 261-290, and 24 (1973). (In Hungarian.)

82. Nevai, G.P., Mean convergence of Lagrange interpolation, I, J. Approximation Theory (to appear).

83. Newman, J. and W. Rudin, Mean convergence of orthogonal series, Proc. Amer. Math. Soc. 3 (1952), 219-222.

84. Nikol'skii, S.M., Inequalities for entire functions of finite type and their applications to the theory of differentiable functions of several variables, Trudy Mat. Inst. Steklov, 38 (1951), 244-278. (In Russian.)

85. Pollard, H., The mean convergence of orthogonal series, I, II,III, Proc. Nat. Acad. Sci. USA 32 (1946), 8-10, Trans. Amer. Math. Soc. 63 (1948), 355-367, Duke Math. J. 16 (1949), 189-191.

86. Potapov, M.K., Estimates for polynomials and their derivatives, Vestnik Moskov. Univ. Ser. I Mat. 2 (1960) 10-20.

87. Potapov, M.K., On approximation by algebraic polynomials in L_p, Investigations on Modern Problems of Constr. Func. Th., edited by V.I. Smirnov, Moscow, 1961, 64-69. (In Russian.)

88. Riesz, M., Eine trigonometrische interpolationsformel und einige Ungleichungen für Polynome, Jahresb.der Deutschen Math. Ver. 23 (1914), 354-368.

89. Shohat, J., Application or orthogonal Tschebycheff polynomials to Lagrangean interpolation and to the general theory of polynomials, Ann. di Math. 18 (1939), 201-238.

90. Stein, M., Interpolation in polyomial classes and Markoff's inequality, Duke Math. J. 24 (1957), 467-476.

91. Szegő, G., Orthogonal Polynomials, AMS Coll. Publ., Providence, Rh. I., 1967.

92. Turán, P., On some problems in the theory of the mechanical quadrature, Mathematics (Cluj) (31)8 (1966), 181-192.

93. Uspensky, J.V., On the convergence of quadrature formulas related to an infinite interval, Trans. Amer. Math. Soc.

26 (1928), 542-559.

94. Vajnerman, Yu. R., Deviation of the Lagrange-Jacobi interpolation process on classes of continuous functions, Doklady AN SSSR 213 (1973), 509-511. (In Russian.)

95. Vidensky, V.S., Extremal estimates of the derivatives of polynomials on intervals, Investigations on Modern Problems of Constr. Func. Th., edited by V.I. Smirnov, Moscow, 1961, 98-106. (In Russian.)

96. Wiener, N., The quadratic variation of a function and its Fourier coefficients, Mass. J. of Math. 3 (1924), 72-94.

97. Wing, G.M., The mean convergence of orthogonal series, Amer. J. Math. 72 (1950), 792-807.

98. Young, L.C., Sur une généralisation de la notion de variation de puissance p-iéme bornée an seus de M. Wiener, et sur la convergence des séries de Fourier, Comptes Rendus Paris 204 (1937), 470-472.

99. Zygmund, A., Trigonometric Series, vol. I, II, Cambridge Univ. Press, Cambridge, 1959.

Paul Nevai* ‡
University of Wisconsin
Department of Mathematics
213 Van Vleck Hall
Madison, Wisconsin 53706

* Supported by the American Mathematical Society.

‡ This paper was presented at the Symposium by Professor Richard Askey of the University of Wisconsin.

FITTING SURFACES TO SCATTERED DATA

Larry L. Schumaker

This paper is a survey of a variety of numerical methods for fitting a function to data given at a set of points scattered throughout a domain in the plane. We discuss four classes of methods: (1) global interpolation, (2) local interpolation, (3) global approximation, and (4) local approximation. We also discuss two-stage methods and contouring. The surfaces constructed will include polynomials, spline functions, and rational functions, among others.

1. Introduction

Our aim is to survey methods for solving the following problem.

PROBLEM 1.1. Let D be a domain in the (x,y)-plane, and suppose F is a real-valued function defined on D. Suppose we are given the values $F_i = F(x_i, y_i)$ of F at some set of points (x_i, y_i) located in D, $i = 1,2,\ldots,N$. Find a function f defined on D which reasonably approximates F.

This problem is, of course, precisely the problem of fitting a surface to given data. In many cases the domain D is a rectangle and the data points lie on a rectangular grid. There are, however, many practical problems (see the following section for some specific examples), where D is of unusual shape and where the data points are irregularly scattered throughout D. Thus, while we shall pay some attention to special methods for regularly spaced data, we are actually more interested in the general case.

There are basically two approaches to handling Problem 1.1. First, we may try to construct a function f which interpolates

Preceding page blank

the data exactly; i.e., such that

$$(1.1) \quad f(x_i, y_i) = F_i, \qquad i = 1,2,\ldots,N.$$

This approach may be desirable when the function values at the data points are known to high precision and where it is highly desirable that these values be preserved by the approximating function.

The second approach involves constructing f which only approximately fits the data. This may be regarded as data smoothing and will be desirable when (as is often the case) the data are subject to inaccurate measurement or even errors. The question of whether interpolation or approximation should be used will not be discussed further here--this is a problem which must be settled for the individual problem at hand.

In discussing Problem 1.1, it will be convenient to make a further distinction between those methods which are local in character (i.e., where the value of the constructed surface f at the point (x,y) depends only on the data at relatively nearby points) and those methods which are global in nature. Thus, we discuss four categories of methods in sections 3-6: (1) global interpolation, (2) local interpolation, (3) global approximation, and (4) local approximation. In each of these sections we further subdivide the material according to the type of functions being used and the type of data (scattered or not) for which the method is suitable.

In discussing methods which apply only to special arrangements of data points, we have two objectives in mind. First, the methods are of interest in their own right. More importantly in terms of Problem 1.1, however, such methods can also be used in two-stage processes in which we first construct a surface g based on the scattered data, and then use g to generate regular data for the construction of another (perhaps smoother or more convenient) surface f. Such two-stage methods

will be discussed (along with several examples) in more detail in section 7.

For many of the methods based on regular data and some of those for scattered data, error bounds are available to indicate how well smooth functions are approximated by the surface constructed. We do not have space to go into the extensive literature on error bounds. A simple test of how well a method will approximate smooth functions is, however, provided by its ability to reproduce polynomial surfaces exactly (that is, if F is a polynomial in x and y up to a certain degree, then the surface f is identically equal to F). For many of the methods we will be able to indicate the corresponding degree of exactness.

In many of the applications of surface-fitting techniques (cf. the examples in section 2), the ultimate aim is to use the data to construct a contour map of the unknown function. Since F is known only at the data points, we must be content to construct a contour map for one of our fitted surfaces. In section 8 we discuss some approaches to accomplishing this numerically.

We close this introduction with a disclaimer--this survey does not include all possible methods for fitting surfaces to scattered data. For example, we have not discussed Fourier series methods, spatial filtering, and other such related statistical techniques. In addition, the set of references for those methods which we have discussed are also not complete. My original intention was to compile as complete a bibliography as possible, but the sheer bulk of relevant papers and my inability to locate all of them convinced me to settle for less. I have opted to quote a fairly representative list of papers, including several other surveys. Further references can be found by consulting these. I shall be very happy to receive information on references and methods I have overlooked.

2. Examples

In this section we shall quote several explicit examples of Problem 1.1 to emphasize the fact that unusually shaped regions and scattered data do arise frequently in practice.

EXAMPLE 2.1. Petroleum exploration. In exploring for petroleum, the contours of various underground layers of sandstone, shale, limestone, etc. can be important indicators of possible oil fields. Frequently, data on such layers is available from exploratory wells, which, however, have most likely been drilled at locations scattered randomly throughout some geographical region of interest. To quote a specific example, Robinson, Charlesworth, and Ellis [166] consider precisely this problem for some data obtained from 7,500 wells drilled in Alberta. For another example of this type, see Whitten and Koelling [208].

Problems similar to that mentioned in Example 2.1 arise frequently in cartography and submarine topography where the measurements represent actual elevations. In some cases the measurements must be taken from photographs or from sonar measurements and are usually subject to some measurement error (eg. see Kubik [125] for a discussion of photogrammetry).

EXAMPLE 2.2. Geological maps. There are a great many problems in Geology and the earth sciences in which the data arises from some other function of location besides actual elevations. For example, some geological variables of interest might include concentrations of various chemicals, specific gravity, electrical resistivity, grain size, texture, optical properties, isotope ratios, etc. To quote a specific example, Bhattacharyya [21, 22] discusses methods for fitting a surface to measurements (taken by airborne sensors) of magnetic potentials over a certain portion of the Yukon. See also Bhattacharyya and Raychaudhuri [23] and Crain and Bhattacharyya [61].

The importance of surface-fitting methods in the earth sciences can be judged by the large number of papers in the area relating to various fitting methods. For a further list of problems and a discussion of some of the methods which have been applied, see the books of Bohrenberg and Giese [31], Chorley [51], David [62], Harbaugh and Merriam [98], and Merriam [140]. Recent survey papers include Whitten [203, 205] and Whitten and Koelling [207]. To add just a few more of the papers in the geological literature dealing with surface fitting to our list, we mention Anderson [7], Grant [91], Hessing, Lee, and Pierce [114], Holroyd and Bhattacharyya [115], Kubik [123, 125], Norcliffe [151], Reilly [162], Whitten [200, 201, 204], and Whitten and Koelling [206].

EXAMPLE 2.3. Heart potentials. In order to diagnose certain abnormal heart conditions, it is desired to make a series of several hundred contour maps of the heart potential field at time steps of 1/100 of a second throughout a heart beat. Data on these heart potentials can be obtained by fitting the patient with a shirt containing probes. Because of body geometry, when this shirt is flattened out it takes the nonrectangular form illustrated in Figure 1. Although the probes could be arranged fairly regularly in this domain, because of the added signifi-

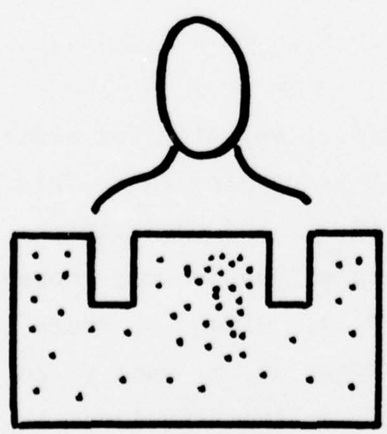


Figure 1. Heart Potential Measurements

cance of frontal measurements, in practice more probes are fitted there than in the back. This example was brought to my attention by Ms. Patrizia Ciarlini of Rome.

Potential fields arise in many other applications. We have already mentioned Geology in Example 2.2. For some examples in modelling plasmas see Buneman [40]. The problem arises in Biersack and Fink [24] in experimentally studying crystal structure using neutron bombardment. Data from waveform distortion in electronic circuits can be found in Akima [5, 6].

3. Global interpolation methods

In this section we outline several methods for solving the interpolation problem (1.1).

3.1 Polynomial interpolation. (Scattered data). The general theory of finite dimensional interpolation is, of course, very well known (e.g., see Davis [63]). Briefly, if $\{\phi_i\}_1^N$ are N functions defined on the domain D, then the function

$$(3.1)\quad f(x,y) = \sum_{j=1}^{N} a_j \phi_j(x,y)$$

will satisfy (1.1) if and only if $\{a_j\}_1^N$ is a solution of the linear system

$$(3.2)\quad \sum_{j=1}^{n} a_j \phi_j(x_i,y_i) = F_i, \quad i = 1,2,\ldots,N.$$

This system has a (unique) solution for arbitrary choices of data precisely when it is nonsingular. This depends on the choice of functions $\{\phi_j\}_1^N$ and the location of the data points.

To illustrate this method, we may choose the $\{\phi_j\}_1^N$ to be polynomials in x and y. Given N, there is some leeway in the choice of which powers of x and y to use. For example, with $N = 3$ one could use the functions $1, x, y$ or possibly the functions $1, x^2, y^2$, etc. When N is of the form $N =$

$(d+1)(d+1)$, we might use the functions

$$\{\phi_j(x,y)\}_1^N = \{x^\nu y^\mu\}_{\nu=0,\mu=0}^{d\ \ ,d}$$

As simple as this sounds, there are some serious difficulties with polynomial interpolation of scattered data. For openers, it is not so easy to guarantee that the system (3.2) is nonsingular. To give a very simple example, consider the case $N = 3$ with the functions $1, x, y$. If the three data points happen to lie on a line, then (3.2) will in fact be singular. Even when (3.2) is nonsingular, it will often be the case (at least if N is moderately large) that the system will be ill-conditioned. Finally, as is well known, polynomials of even moderate degree exhibit a considerable oscillatory character, and the resulting surface (even though it is C^∞) is often too undulating to be acceptable. The general problem of polynomial interpolation to scattered data is not usually treated in Numerical Analysis and Approximation Theory books (see, however, Kunz [126], Prenter [157], and Steffenson [186]). Some papers dealing with the question include Guenther [93], Thatcher [189], Thatcher and Milne [190], and Whaples [197]. Assuming the interpolant exists, error bounds have been studied in Ciarlet and Raviart [52-55].

Let

$$(3.3)\qquad \mathcal{P}_{m,n} = \text{span}\ \{x^\nu y^\mu\}_{\nu=0,\mu=0}^{m\ \ ,n}$$

be the space of polynomials of degree m in x and of degree n in y. This linear space is of dimension $(m+1)(n+1)$ and is, in fact, the tensor product of the linear spaces $\mathcal{P}_m$ and $\mathcal{P}_n$. It is perhaps of interest to note that there always exists a (usually nonunique) polynomial $p \in \mathcal{P}_{N,N}$ which solves the interpolation problem (1.1), no matter how the data points are positioned, see Prenter [158].

3.2 Polynomial interpolation (gridded data). We begin this subsection by defining what we mean by gridded data. Let

$$(3.4)\qquad H = [a,b] \times [c,d]$$

be a rectangle, and let

$$(3.5)\qquad \begin{aligned} a &= x_0 < x_1 < \ldots\ldots < x_{k+1} = b \\ c &= y_0 < y_1 < \ldots\ldots < y_{\ell+1} = d. \end{aligned}$$

We suppose now that F is a function defined on H, and that we have the values of F at the corner points of the rectangular grid defined by (3.5); i.e.,

$$(3.6)\qquad F_{ij} = F(x_i, y_j), \quad \begin{aligned} i &= 0,1,\ldots,k+1 \\ j &= 0,1,\ldots,\ell+1. \end{aligned}$$

This is a total of $N = (k+2)(\ell+2)$ data points.

It is quite easy to show that there exists a unique polynomial p in the class $\mathcal{P}_{k+1,\ell+1}$ (cf. the definition (3.3)) which interpolates the gridded data given in (3.4)-(3.6). In fact, p can be written down explicitly in terms of the one-dimensional Lagrange polynomials as

$$(3.7)\qquad p(x,y) = \sum_{i=0}^{k+1} \sum_{j=0}^{\ell+1} F_{ij} L_i(x) \tilde{L}_j(y),$$

where the $\{L_i(x)\}_0^{k+1}$ and $\{\tilde{L}_j(y)\}_0^{\ell+1}$ are the usual one-dimensional Lagrange polynomials associated with the interpolation points $\{x_i\}_0^{k+1}$ and $\{y_j\}_0^{k+1}$, respectively. Interpolation of gridded data by polynomials has been discussed in various books and papers--we do not bother with a long list. See e.g. Prenter [157] or Steffenson [186]. More recently, there has been considerable work on Hermite and osculatory interpolation in several variables; see e.g. Ahlin [3], Haussman [99,101,102], and Salzer [168-170].

3.3 Shepard's method. In this subsection we discuss a method of Shepard [180] and some modifications of it. The method applies to arbitrarily spaced data, and the interpolating function can be written down explicitly.

Let ρ be some metric in the plane, for example the usual distance metric. Given a point (x,y), let $r_i = \rho((x,y),(x_i,y_i))$ for $i = 1,2,\ldots,N$. Let $0 < \mu < \infty$. Then Shepard's interpolation formula is defined by

$$(3.8)\quad f(x,y) = \begin{cases} \left(\sum_{i=1}^{N} \frac{F_i}{r_i^{\mu}}\right) \Big/ \left(\sum_{i=1}^{N} \frac{1}{r_i^{\mu}}\right), & \text{when } r_i \neq 0, \text{ all } i \\ F_i, & \text{when } r_i = 0. \end{cases}$$

The formula (3.8) is defined for all points (x,y) in the plane R^2. It is clear from the definition that it interpolates the values F_i at the data points (x_i,y_i), $i = 1,2,\ldots,N$. The value of $f(x,y)$ at nondata points is obtained as a weighted average of all the data values, where the i^{th} measurement is weighted according to the distance of (x,y) from the point (x_i,y_i).

We shall briefly recount some of the properties of Shepard's formula. First, by converting all of the terms to a common denominator, it can be shown that

$$(3.9)\quad f(x,y) = \sum_{i=1}^{N} F_i A_i(x,y),$$

where

$$(3.10)\quad A_i(x,y) = \frac{\prod_{\substack{j=1 \\ j\neq i}}^{N} [r_j(x,y)]^{\mu}}{\sum_{k=1}^{N} \sum_{\substack{\ell=1 \\ \ell\neq k}}^{N} [r_\ell(x,y)]^{\mu}}, \quad i = 1,2,\ldots,N .$$

These functions satisfy

(3.11) $A_i(x_j, y_j) = \delta_{ij}, \qquad i,j = 1,2,\ldots,N.$

The representation (3.9) is numerically more stable than the original formula (3.8).

In view of its definition, we see that the function $f(x,y)$ constructed by Shepard is not a simple polynomial or rational function. It is clear, however, that except for the points (x_i, y_i), it is analytic everywhere in the plane. Its behavior in the vicinity of the data points (x_i, y_i) depends on the size of μ. It can be shown that for $0 < \mu \leq 1$, f has cusps at these points. For $1 < \mu$, f has flat spots at the data points (i.e., the partial derivatives vanish there). We also observe the interesting property that

(3.12) $$\min_{1\leq i\leq N} F_i \leq f(x,y) \leq \max_{1\leq i\leq N} F_i.$$

We may also note that if the data came from a constant function, i.e., $F_i = c$, $i = 1,2,\ldots,N$, then f is also the constant function $f \equiv c$.

We now comment on the choice of μ. To get smooth surfaces without cusps, it is desirable to take $1 < \mu$. On the other hand, if μ is relatively large, then the surface tends to become very flat near the data points and consequently quite steep at points in between. Experiments (cf. Gordon and Wixom [90], Poeppelmeir [155], and Shepard [180]) seem to indicate that a choice of $\mu = 2$ is perhaps a good tradeoff. ([155] contains several examples showing the behavior as a function of μ.)

There are several drawbacks to Shepard's method (3.8), as pointed out by Shepard [180] himself. First, if N is large, then there is a very considerable amount of calculation involved in evaluating $f(x,y)$ at a particular point. Secondly, the weights are assigned on the basis of the distance of points from (x,y) only, not their direction. Finally, the flat spots

in the neighborhood of the data points is somewhat disturbing. The first of these objections can be met by defining a local version of the formula, which we shall do in section 4.5. It is possible to construct an analogous formula which accounts for direction. For details, see Shepard [180]. Finally, we briefly discuss handling the flat spots.

Suppose in addition to the function values F_i at each point (x_i, y_i) we also have estimates FX_i and FY_i of $F_x(x_i, y_i)$ and $F_y(x_i, y_i)$. Then we may consider the function

$$(3.13)\quad f(x,y) = \sum_{i=1}^{N} A_i(x,y)\,[F_i + (x-x_i)FX_i + (y-y_i)FY_i].$$

It is easily checked that this function also interpolates, and that

$$(3.14)\quad f_x(x_i,y_i) = FX_i, \quad f_y(x_i,y_i) = FY_i, \quad i = 1,2,\ldots,N.$$

This property may be expressed in the assertion that if the data F_i, FX_i, FY_i came from a plane surface F, then f will exactly reproduce this surface. To use formula (3.13) in practice on the data-fitting Problem 1.1, we have to carry out a two-stage approximation process in which the first stage consists of some method for estimating the slope at each of the data points.

It might be of practical interest in some cases to construct still a more sophisticated version of Shepard's formula which would exactly reproduce higher-order polynomial surfaces. One approach to doing this is to use the following lemma.

LEMMA 3.1. (Barnhill [15]). Let P and Q be linear projections of some linear space of functions $\mathcal{F}$ into itself. Suppose that Q exactly reproduces the linear subspace $E \subset \mathcal{F}$; i.e.,

$$(3.15)\quad Qp = p, \quad \text{all } p \in E.$$

In addition, suppose that $\{\lambda_i\}_1^m$ is a set of linear functionals on $\mathcal{F}$, and that

(3.16) $\lambda_i Pf = \lambda_i f$, all $f \in \mathcal{F}$, $i = 1,2,\ldots,m$.

Then the Boolean sum projector

(3.17) $P \oplus Q = P + Q - PQ$

enjoys the function precision of Q (i.e., reproduces E) and the interpolation properties of P (i.e., (3.16) also holds for $P \oplus Q$).

This result permits the construction of interpolation schemes using Shepard's formula which reproduce higher-order surfaces. For an example, see Poeppelmeir [155] where Shepard's formula is combined with a certain local interpolation scheme which reproduces quadratic surfaces. In closing this section we note that Shepard's formula can also be interpreted as arising from weighted least squares--see section 5.1.

3.4 Spline interpolation (scattered data). Suppose X is a linear space of "smooth" functions defined on the domain D, and let

(3.18) $U = \{f \in X: \ f(x_i,y_i) = F_i, \ \ i = 1,2,\ldots,N\}$.

U is the set of smooth functions which interpolate. Now suppose that Θ is a functional on X which measures the smoothness of an element in X--the smaller $\Theta(f)$ is, the smoother f is. Then we may consider the following minimization problem:

(3.19) Find $s \in U$ such that $\Theta(s) = \inf_{u \in U} \Theta(u)$.

The function s will be the smoothest interpolant, and in view of the similarity with classical spline approximation, s is called a spline function interpolating F. The basic questions concerning spline interpolation center around existence, uniqueness, characterization, and construction. A quite general

abstract theory of spline interpolation has been built up (see eg. Laurent [127] and references therein). In this section we quote some specific examples which can be used on Problem 1.1.

Where X is a semi-Hilbert space, $\Theta(f) = \|f\|$, where $\|\cdot\|$ is a seminorm on X, and $N = \{f \in X: \|f\| = 0\}$, it is possible to show (under some additional mild conditions on X, see Duchon [72, 73]) that problem (3.19) always has a solution which is unique up to an element in N. Moreover, it can be shown that there exists a reproducing kernel K defined on $D \times D$ such that

$$(3.20) \quad s(x,y) = \sum_{i=1}^{N} a_i K((x,y);(x_i,y_i)) + \sum_{i=1}^{d} b_i p_i(x,y),$$

where $\{p_i\}_1^d$ is a basis for N. Moreover, the coefficients $\{a_i\}$ and $\{b_i\}$ can be determined from the linear system of equations

$$(3.21) \quad \sum_{j=1}^{N} K((x_j,y_j);(x_i,y_i))a_i + \sum_{i=1}^{d} b_i p_i(x_j,y_j) = F_j, \quad j=1,\ldots,N$$

$$\sum_{i=1}^{N} a_i p_k(x_i,y_i) = 0, \quad k = 1,2,\ldots,d.$$

The development with semi-Hilbert spaces in Duchon [72, 73] is an extension of earlier work of Atteia [10-12] and Thomann [192-193] using Hilbert spaces. The essential difficulty in applying the general results is the construction of an appropriate reproducing kernel. We turn now to some specific examples.

Suppose X is the space of all functions on the rectangle $D = H$ (cf. (3.4)) which have (distributional) derivatives up to order 2 which lie in $L^2(H)$. For $f \in X$, let

$$(3.22) \quad \Theta(f) = \iint_D |D_x^2 f|^2 + 2|D_x D_y f|^2 + |D_y^2 f|^2.$$

The reproducing kernel in this case can be written down as an infinite series involving sin and cos, and the space N is spanned by 1, x, and y. Similarly, if we replace H by the unit disc UD, the kernel can be computed as an infinite series (see Atteia [10-12] and Thomann [192-193]). Thomann considers computation of these splines by approximating the infinite series--FORTRAN programs are also included.

If we replace the bounded sets H or UD by the entire plane R^2 and introduce an appropriate space X, it is possible to obtain explicit expressions for the reproducing kernel. This is the content of Duchon [72, 73]. In particular, let $\tilde{H}^s$ be the set of all tempered distributions f on R^2 whose Fourier transforms $\hat{f}$ satisfy $\int |\hat{f}| t^{2s} dt < \infty$. Let X^{ms} denote the set of all functions which have derivatives up to order m lying in $\tilde{H}^s$. Our first example concerns the space X^{20}. If we choose Θ as in (3.22), then the interpolating spline solution of (3.19) is of the form

$$(3.23) \quad s(x,y) = \sum_{i=1}^{N} a_i r_i^2(x,y) \log (r_i(x,y)) + b_1 x + b_2 y + b_3,$$

where $r_i(x,y) = [(x-x_i)^2 + (y-y_i)^2]^{\frac{1}{2}}$. The coefficients are determined from the system (3.21) with $d = 3$, $N = \text{span } \{1,x,y\}$, and $K(z,w) = |z-w|^2 \log(z-w|$. Duchon refers to this type of spline as a thin plate spline since the expression Θ relates to the energy in a thin plate forced to interpolate the data. This spline belongs to $C(R^2)$.

As a second example, suppose we consider $X = X^{21}$. In this case the solution of (3.19) with Θ given by (3.22) has the form

$$(3.24) \quad s(x,y) = \sum_{i=1}^{N} a_i (r_i(x,y))^3 + b_1 x + b_2 y + b_3.$$

Here $K(z,w) = |z-w|^3$. Duchon [72, 73] refers to these splines

as <u>pseudo-cubic splines</u> because of the analogy with the cubic splines in one variable. They belong to $C^1(R)$. <u>Pseudo quintic splines</u> etc. are also considered in Duchon [72, 73].

A similar program has been carried out by Mansfield [133-137] for some spaces of smooth functions defined on a rectangle H. In [136] she considers a space of functions $T^{m,n}(\alpha,\beta)$, where m and n are positive integers and $a \leq \alpha \leq b$, $c \leq \beta \leq d$. This space is actually defined by completion of a set of tensor product functions with respect to an appropriate inner-product, and we do not want to define it precisely here. A function $f \in T^{m,n}(\alpha,\beta)$ has the following properties, however:

$$(3.25)\quad \begin{cases} f^{(i,j)} \in C(H), \quad i = 0,1,\ldots,m-1 \text{ and } j = 0,1,\ldots,n-1 \\ f^{(s-j-1,j)}(x,\beta) \in AC[a,b] \text{ and } f^{(s-j,j)}(x,\beta) \in L^2[a,b], \\ \qquad\qquad j = 0,1,\ldots,n-1 \\ f^{(i,s-i-1)}(\alpha,y) \in AC[c,d] \text{ and } f^{(i,s-i)}(\alpha,y) \in L^2[c,d], \\ \qquad\qquad i = 0,1,\ldots,m-1 \\ f^{(m-1,n-1)} \in AC(H) \text{ and } f^{(m,n)} \in L^2(H), \end{cases}$$

where AC stands for the space of absolutely continuous functions and where $s = m+n$. By constructing an appropriate reproducing kernel, she is able to solve problem (3.19) with

$$(3.26)\quad \Theta(f) = \iint_H [f^{(m,n)}]^2 + \sum_{j=0}^{n-1} \int_a^b [f^{(s-j,j)}(x,\beta)]^2 dx + \sum_{i=0}^{m-1} \int_c^d [f^{(i,s-i)}(\alpha,y)]^2 dy.$$

In [133], Mansfield carries out a similar analysis for a space of functions $R^{m,n}$ defined on the rectangle H. Here $R^{m,n} = L_2^m[a,b] \times L_2^n[c,d]$, where $L_2^m[a,b]$ is the usual Sobolev space of functions with absolutely continuous derivatives up to order $m-1$, and with $f^{(m)} \in L^2[a,b]$. By constructing an

appropriate reproducing kernel, she now solves problem (3.19) with

$$\Theta(f) = \iint_H [f^{(m,n)}]^2 + \sum_{j=0}^{n-1} \int_a^b [f^{(m,j)}(x,c)]^2 dx + \sum_{i=0}^{m-1} \int_c^d [f^{(i,n)}(a,y)]^2 dy \,. \tag{3.27}$$

The solution turns out to be a piecewise polynomial of degree $2m-1$ in x and of degree $2n-1$ in y. It is also in $C^{2m-2,2n-2}(H)$. For the particular case of gridded data, it reduces to the tensor product of one-variable splines (cf. the following section). Other more general definitions of Θ are also considered (with minor modifications on the one-dimensional integrals).

A more algebraic approach to constructing multidimensional spline functions (which also involves certain kernel functions) has been taken by Schaback [173-174]. His two-dimensional kernel function is obtained as a tensor product of one-dimensional kernels.

3.5. Spline interpolation (gridded data). The problem of constructing interpolating splines in two dimensions with gridded data as in (3.4)-(3.6) is, of course, a special case of the general problems discussed in subsection 3.4. The development of the gridded data case predated the more general development and, moreover, is considerably simpler. There are a great many papers on two-dimensional polynomial splines and generalizations. We do not have space here to discuss all of them in detail. We shall be content to quote some of the papers and to give a somewhat more complete discussion of polynomial splines, which are the most widely used splines for this problem.

Some early papers dealing with two-dimensional interpolating splines include Birkhoff and de Boor [26], Birkhoff and

Garabedian [27], Price and Simonson [159], and Theilheimer and Starkweather [191]. In [26] certain bicubic splines were introduced which were later studied in detail in de Boor [32]. The problem was to minimize

$$(3.28)\quad \int_a^b \int_c^d [f^{(2,2)}(x,y)]^2 dxdy$$

over all appropriately smooth functions on the rectangle H which interpolate the gridded data (3.4)-(3.6). It was found that the solution of this problem was a certain bicubic function with global smoothness $C^2(H)$. This problem was generalized to minimizing

$$(3.29)\quad \Theta(f) = \int_a^b \int_c^d [f^{(m,n)}(x,y)]^2 dx\,dy, \quad m = 2p,\ n = 2q$$

in Ahlberg, Nilson and Walsh [1,2], whose solution involves certain higher-order polynomial splines. Since they are widely used, we give a short algebraic treatment here.

The points $\{x_i\}_0^{k+1}$ and $\{y_j\}_0^{\ell+1}$ define a partition of the intervals $[a,b]$ and $[c,d]$, respectively (cf. (3.5)). Suppose now that $x_{1-m} \le \dots \le x_{-1} \le a < b \le x_{k+2} \le \dots \le x_{k+m-1}$ and $y_{1-n} \le \dots \le y_{-1} \le c < d \le y_{\ell+2} \le \dots \le y_{\ell+n-1}$ are chosen arbitrarily. Let $\{N_i^m\}_{1-m}^k$ be the B-splines associated with the x-partition, and let the B-splines associated with the y-partition be denoted by $\{N_j^n(y)\}_{1-n}^{\ell}$. For a complete discussion of B-splines and their properties, see de Boor [36] in this volume (or [33]). Let

$$(3.30)\quad N_{ij}(x,y) = N_i^m(x)N_j^n(y), \quad i = 1-m,\dots,k \text{ and } j = 1-n,\dots,\ell.$$

The linear space

$$(3.31)\quad \mathcal{S} = \operatorname{span}\{N_{ij}(x,y)\}_{i=1-m,\ j=1-n}^{k\ \ ,\ \ \ell}$$

is clearly of dimension $(k+m)(\ell+n)$. We may now construct an

element in (3.31) which interpolates to the gridded data.

Since there are only $(k+2)(\ell+2)$ data points on the grid (cf. (3.4)-(3.6), it is clear that if we use $\mathcal{S}$ to interpolate, we have

$$(3.32)\qquad (k+m)(\ell+n)-(k+2)(\ell+2) = (k+2)(n-2)+(\ell+2)(m-2)+(n-2)(m-2)$$

free parameters. Thus, to uniquely define a spline, one must add additional conditions. Recall that $m=2p$ and $n=2q$. Then we might add the extra conditions

$$(3.33)\qquad \begin{aligned} s^{(\nu,0)}(x_0,y_j) &= s^{(\nu,0)}(x_{k+1},y_j)=0, && j=0,1,\ldots,\ell+1 \\ & && \nu=p,\ldots,m-2 \\ s^{(0,\mu)}(x_i,y_0) &= s^{(0,\mu)}(x_i,y_{\ell+1})=0, && i=0,1,\ldots,k+1 \\ & && \mu=q,\ldots,n-2 \end{aligned}$$

and

$$(3.34)\qquad \begin{aligned} s^{(\nu,\mu)}(x_0,y_0) &= s^{(\nu,\mu)}(x_0,y_{\ell+1}) = s^{(\nu,\mu)}(x_{k+1},y_0) \\ &= s^{(\nu,\mu)}(x_{k+1},y_{\ell+1})=0, \quad \nu=p,\ldots,m-1 \\ &\qquad\qquad\qquad\qquad\qquad \mu=q,\ldots,n-1. \end{aligned}$$

These are called the natural boundary conditions, and it can be shown that the system of equations

$$(3.35)\qquad \sum_{i=1-m}^{k}\ \sum_{j=1-n}^{\ell} a_{ij}N_{ij}(x_\alpha,y_\beta) = F_{\alpha\beta}, \quad \begin{aligned}\alpha&=0,1,\ldots,k+1\\ \beta&=0,1,\ldots,\ell+1\end{aligned}$$

coupled with the conditions (3.33)-(3.34) provides a nonsingular system of equations for the coefficients $\{a_{ij}\}$. This system has convenient bandedness properties if the equations are arranged properly. The resulting spline is precisely the solution of the minimization problem (3.29). The boundary conditions (3.33)-(3.34) are the natural ones associated with the problem (3.29). However, it is also possible to specify lower-order derivative information along the boundary and also obtain a nonsingular system of equations. The resulting spline, called Type I, can also be shown to satisfy an appropriate minimization

problem. However, for data-fitting purposes, to use the interpolant with boundary derivative data one would first have to perform a first-stage approximation to find estimates for the required derivatives.

The best-known case of the above spline interpolation is the case $m = n = 4$, i.e., bicubic spline interpolation. In this case the surface constructed is a piecewise bicubic with global smoothness $C^2(H)$. The natural boundary conditions set second-derivative values to 0. Programs for computing natural bicubic interpolating splines can be found in the IMSL Library [117] in FORTRAN. FORTRAN programs for Type I bicubic splines can be found in Koelling and Whitten [121], where the required boundary information is assumed to be input. An ALGOL program for computing Type I bicubic splines in which boundary data are automatically computed by fitting one-dimensional splines appears in Späth [183].

Bicubic spline interpolation has been widely applied. For some references in the Geology literature, see eg. Anderson [7], Bhattacharyya [22], Holroyd and Bhattacharyya [115], Koelling and Whitten [121], and Whitten and Koelling [206].

Problem (3.29) has been widely generalized in the spline literature. Instead of minimizing ordinary derivatives, one may introduce general linear operators, and instead of dealing with point evaluation functionals, more general linear functionals may be permitted. To list some (but by no means all) papers dealing with such generalizations, we mention Arthur [8,9], Birkhoff, Schultz and Varga [29], de Boor [34], Delvos [65,66], Delvos and Schempp [68,69], Delvos and Schlosser [70], Fisher and Jerome [78,79], Haussmann [100], Haussmann and Munch [104], Munteanu [143,144], Nielson [148,150], Ritter [164,165], Sard [171,172], Schoenberg [176], Schultz [177,178], Späth [184,185], and Zavialov [209-212]. On L-shaped regions and other polygons

see Birkhoff [25] and Carlson and Hall [44-49].

We close this section by mentioning another direction of generalization which has led to a considerable development, the idea of spline blending. These methods are useful for construction of a surface which interpolates not only function values at isolated points but on the grid lines themselves; i.e.,

$$\begin{aligned} &(3.36) \quad f(x,y_j) = F(x,y_j) \qquad a \le x \le b \quad \text{and} \quad j = 0,1,\ldots,\ell+1 \\ &\qquad\qquad f(x_i,y) = F(x_i,y) \qquad c \le y \le d \quad \text{and} \quad i = 0,1,\ldots,k+1. \end{aligned}$$

To use such blending methods one must have F defined on the grid lines. Thus, the methods could be of value as second-stage processes. We do not have space to go into detail on spline-blended methods. We refer to the recent book of Barnhill and Riesenfeld [20] for a collection of papers on the subject and for further references. See also the papers of Gordon [84-87] and Gordon and Hall [88]. Recently, considerable effort has gone into showing that blending methods also arise as solutions of appropriate variational problems; see the papers of Delvos [65], Delvos and Kosters [66], and Delvos and Malinka [67].

4. Local interpolation methods

The interpolation methods discussed in section 3 were global in nature--that is, the value $f(x,y)$ of the constructed surface at any given point (x,y) in D depends on the values of all of the data points. This generally means that to compute a representation for f one has to solve a fairly large system of equations, and to evaluate $f(x,y)$ one generally has to carry out a considerable amount of arithmetic. In this section we shall consider local schemes where the surface depends only on nearby data points. Then the construction will usually lead to (a possibly large number) of <u>small</u> systems of equations, and moreover, the evaluation of the surface at a given point will

usually involve very little computation.

Many of the schemes mentioned in section 3 can be made local in nature by the following simple approach. Suppose that the domain D is partitioned into subdomains: $D = \bigcup_{i=1}^{d} D_i$. We then seek a surface in the form

$$(4.1) \quad f(x,y) = \{f_i(x,y), \quad (x,y) \in D_i, \quad i = 1,2,\ldots,d.$$

To construct each individual f_i, we suppose that $\tilde{D}_i$ are domains containing D_i, which contain only points which are "near" D_i. Then we use the data (and only the data) in $\tilde{D}_i$ to construct f_i. Usually, we can choose $\tilde{D}_i = D_i$. In most cases the most convenient choices for the subdomains D_i are triangles and rectangles. We discuss these two cases first.

4.1. Triangular subregions (scattered data). Suppose that we are given data at points $P_i = (x_i, y_i)$, $i = 1,2,\ldots,N$ scattered throughout the plane, and let D be the convex hull of these points. It is more or less clear that by drawing lines from point to point we can construct a set of triangles with vertices at the P_i which partition D. It is also clear that given any set of points, this triangularization of D is not usually uniquely defined (see Figure 2 below for two different triangularizations of the same region). Moreover, as the figure shows, some triangularizations are superior to others in the sense that they exhibit fewer of the less desirable long thin triangles.

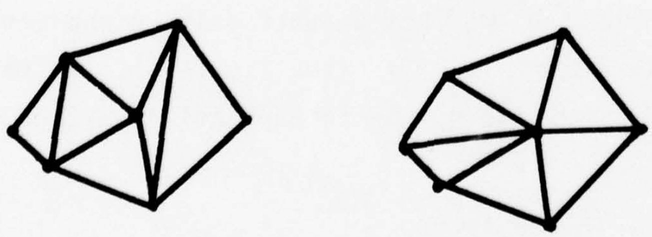


Figure 2. Triangularization

The design of an algorithm to divide a region into acceptable triangles with vertices at prescribed points is not as easy as it sounds. Two algorithms in the literature which are designed to give good triangularizations can be found in Cavendish [50] and in Lawson [128].

The simplest approach to defining a local interpolating surface is to construct $f_i(x,y)$ to be of the form $a_1 + a_2x + a_3y$ in each triangle. The data at the three corners of the triangle determine the coefficients for that piece of f (the corresponding system will be nonsingular provided the triangle is nondegenerate). This procedure produces a piecewise linear surface which, in fact, will be globally continuous. This last property follows from the fact that along the sides of the triangle the functions reduce to straight lines joining the vertices. This method has been used by several authors for data fitting, e.g., Lawson [128] and Whitten [206]. For some contouring routines based on this local interpolation scheme, see section 8.

If we desire to interpolate several sets of data defined on the same triangularization, it may be more convenient to compute Lagrangian functions rather than to compute the surface in each triangle separately. In particular, it is clear that we can construct functions $\{\phi_j(x,y)\}_1^N$ with the property

$$(4.2) \quad \phi_j(x_i,y_i) = \delta_{ij}, \quad i,j = 1,2,\ldots,N.$$

These functions can be constructed as pyramids in such a way that the function ϕ_j has support only on the triangles surrounding the point (x_j,y_j) (see Figure 3). In terms of these Lagrangian functions, the interpolating surface is given by

$$(4.3) \quad f(x,y) = \sum_{j=1}^{N} F_j\phi_j(x,y).$$

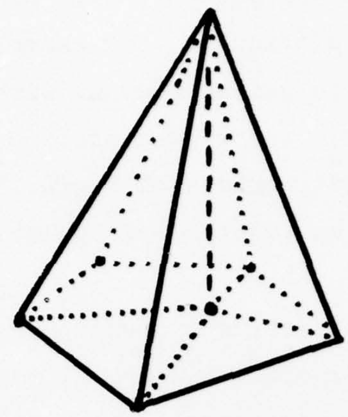


Figure 3. A Lagrange Element

The Lagrangian approach to local interpolation is very reminiscent of the finite element method in which the solution of an operator equation is sought in the form of a linear combination of a set of functions (called elements) with the property (4.2). (See e.g., Prenter [157], Schultz [179], or Strang and Fix [188].) There is no need to restrict the elements to be piecewise linear functions--we may use higher-order polynomials, rational functions, or even more complicated functions. In fact, if we are careful in the construction, we may be able to construct elements with small support but higher global smoothness.

There are a great many papers in the finite-element literature concerned with defining convenient smooth elements (Lagrangian functions with small support). To mention a few, see Barnhill, Birkhoff, and Gordon [16], Barnhill and Gregory [17, 18], Barnhill and Mansfield [19], Birkhoff and Mansfield [28], Bramble and Zlamal [39], Goel [83], Hall [94], Mitchell [141], Mitchell and Phillips [142], Nicolaidis [146,147], Zenisek [213], Zienkowicz [214], and Zlamal [215-217]. The books on finite elements of Aziz [13], de Boor [35], Strang and Fix [188], and Whiteman [198] should also be consulted.

The construction of elements with higher-order smoothness becomes increasingly difficult. For example, it is shown in Mansfield [137] that to get an element with support on the triangles surrounding P_j and with global continuity $C^1(D)$, it is necessary to use polynomials of degree 5 at least. (Matters are somewhat simpler on regular triangularizations, see subsection 4.2 below.)

We close this subsection by mentioning that it is also possible to perform interpolation using elements based on triangles to data which also involves derivatives, or in analogy with the blending methods, to data which includes function values along the edges of the triangles. (See e.g., Barnhill, Birkhoff, and Gordon [16], or Barnhill and Gregory [17,18].) These methods are not directly applicable to the scattered data Problem 1.1, but may be useful as second-stage methods.

4.2. Regular triangularizations. When the data is distributed such that the region can be triangulated into a set of congruent triangles, then it is extremely advantageous to use the Lagrange approach. In particular, in this case we can find an element ϕ with value 1 at (0,0) such that all other elements are translates of ϕ. In this case, f takes the form

$$(4.4)\qquad f(x,y) = \sum_{j=1}^{N} F_j \phi((x,y) - (x_j, y_j)).$$

We illustrate this with a couple of examples. Suppose that we are given data at points chosen from the collection

$$(4.5)\qquad \Omega_1 = \{(i,j)\}_{i,j\in Z} \cup \{(i+\tfrac{1}{2}, j+\tfrac{1}{2})\}_{i,j\in Z}, \quad Z = \{\text{integers}\}.$$

These points lie on the corners of a triangular grid as shown in Figure 4.

It is shown in Zwart [218,p.673] that there exists a function $\phi \in C^1(R^2)$ which is 1 at the origin and 0 at all other points in Ω, and has support on the shaded region in Figure 4.

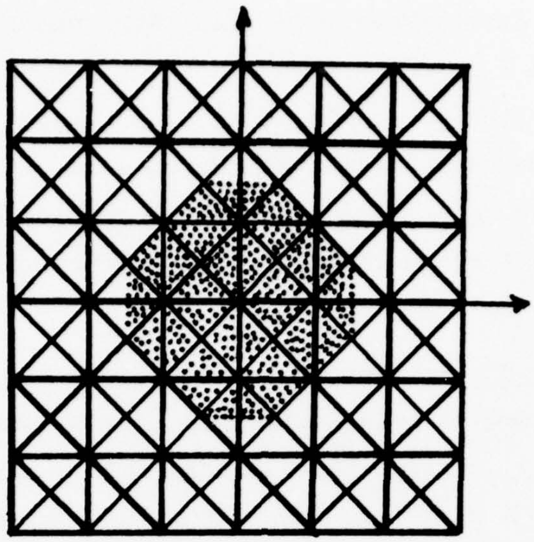


Figure 4. A Regular Triangularization

This function is constructed as a piecewise quadratic polynomial. A similar element has been constructed by Powell [156] (the figure on page 267 of [156] should be rotated 45° to see this).

To give another example, suppose that we consider the set of points Ω_2 which lie at the vertices of the grid defined by equilateral triangles shown in Figure 5.

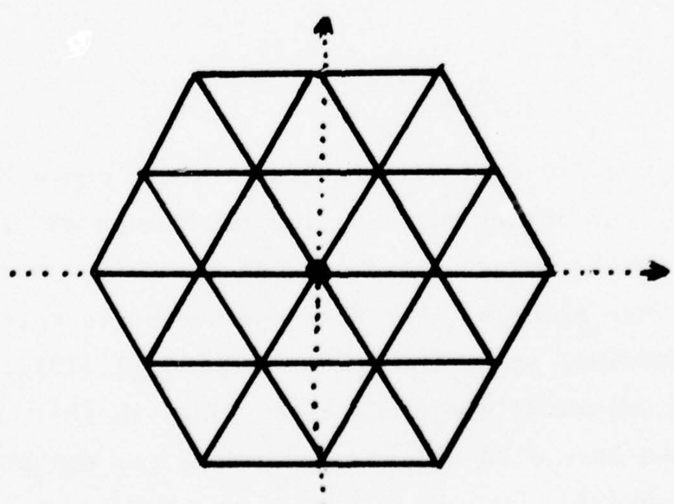


Figure 5. Another Regular Triangularization

It is shown in Fredrickson [81] that there exists a function ϕ which has value 1 at the origin and value 0 at all other points in Ω_2. The function ϕ is in $C^2(R^2)$, consists of piecewise quartics, and has support in the region shown in Figure 5. Fredrickson also constructs a piecewise cubic element with the same support but which is only $C^1(R^2)$. For right triangles see Carlson and Hall [44].

4.3. Rectangular subregions. In this section we suppose that we have data given at points lying on a rectangular grid as in (3.4)-(3.6), and consider local interpolation methods. The simplest approach here (cf. the triangularization case) is to construct a separate bilinear function $f(x,y) = a_1 + a_2x + a_3y + a_4xy$ in each subrectangle, $H_{ij} = [x_i,x_{i+1}] \times [y_j,y_{j+1}]$, using the four corner values to determine the coefficients. Since the bilinear patches reduce to linear functions on the grid lines, the global surface is $C(R)$.

Several authors have considered constructing functions on each of the H_{ij} using higher-order polynomials. This requires additional information in addition to the four corner values. For example, if one seeks a bicubic

$$(4.6) \quad f(x,y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij}x^iy^j,$$

there are 16 coefficients to determine. These could be determined by the four corner values, plus the values of f_x, f_y, and f_{xy} at each corner. To determine these, one must perform some first-stage process. For some approaches to this, see Akima [5], Hessing, et al [114], and Shu, et al [181]. A FORTRAN program for Akima's method can be found in [6]. Nonpolynomial patches have also been considered; e.g., see Birkhoff and Garabedian [27].

The Lagrange (finite element) approach can also be used in

the case of rectangular gridded data. In particular, if we can construct a function satisfying (4.2) with local support, then the surface f given by (4.3) will interpolate and the method will be local in character. As before, the Lagrange approach is especially convenient if the grid is regular, i.e., if all subrectangles H_{ij} are congruent. To illustrate this, suppose that the H_{ij} are actually the unit squares; i.e., the data points lie in the set

$$(4.7) \quad \Omega_3 = \{(i,j)\} \quad i,j \in Z, \quad Z = \{\text{integers}\}.$$

To get a quadratic C^1 element, we may simply rotate the element of Zwart [218] considered in the last section by 45 degrees (cf. Figure 4), or we may take the element of Powell [156].

4.4. <u>Parametric representations</u>. The methods discussed in the last section is concerned with data given on a rectangular grid. By using parametric representations, it is possible to construct similar local interpolating surfaces for data given at the corners of any partition of D consisting of quadrilaterals. In this section we briefly describe how this might proceed.

Suppose Q is a particular quadrilateral subregion of D of interest. In addition, suppose that $x(s,t)$, $y(s,t)$, and $z(s,t)$ are functions defined on the unit square $U = [0,1] \times [0,1]$ with the properties that as (s,t) runs over the boundary of U, $(x(s,t),y(s,t))$ runs over the boundary of the quadrilateral; the four corners of U correspond to the four corners of Q; and $z(s,t)$ takes on the desired data values at the four corners of U. In this case, the triple $(x(s,t),y(s,t),z(s,t))$ provides a parametric representation of a piece of surface defined over Q interpolating the data.

The problem of constructing parametric representations of interpolating functions has been considered in a number of papers. Several papers on these methods and a host of references can be found in the book of Barnhill and Riesenfeld [20]; see

also the survey paper of Shu et al. [181]. Such surfaces are sometimes called Coon's surfaces, cf. Coons [59], and are of considerable interest in the field of computer-aided geometric design. To mention just a few of the actual papers, see Ahuja and Coons [4], Earnshaw and Youille [74], Ferguson [77], Hayes [107], Hosaka [116], and Mangeron [132].

There also has been some effort directed towards constructing elements (Lagrange functions) associated with other less regular subsets of the plane. We mention, for example, the work of Ciarlet and Raviart [55], Wachspress [194,195], and Zlamal [217] in which elements are constructed for domains involving curved edges.

4.5. Local Shepard methods. It is possible to modify the method discussed in subsection 3.3 to make it local. For example, following Shepard [180], suppose we fix $0 < R$ and define

$$(4.8) \qquad \psi(r) = \begin{cases} 1/r & 0 < r \le \frac{R}{3}, \\ \frac{27}{4R}\left(\frac{r}{R} - 1\right)^2, & R/3 < r \le R, \\ 0 \qquad , & R < r \quad . \end{cases}$$

This function is continuously differentiable and vanishes identically for $r > R$. Now with r_i as in (3.8), we define

$$(4.9) \qquad f(x,y) = \begin{cases} \dfrac{\sum_{i=1}^{N} F_i [\psi(r_i)]^{\mu}}{\sum_{i=1}^{N} [\psi(r_i)]^{\mu}}, & \text{when } r_i \ne 0, \text{ all } i \\ F_i \qquad , & \text{when } r_i = 0. \end{cases}$$

Formula (4.9) is defined at all (x,y) in the plane R^2. By definition it interpolates the values F_i at the data points (x_i, y_i), $i = 1,2,\ldots,N$. The values at non-data points are obtained as weighted averages of the data values F_i, but

only those which lie at points within a distance of R of (x,y). Thus, the formula is local.

To use this method in practice it is necessary to choose a reasonable value for R. The aim is to find R so that for every (x,y) a reasonable number of data points will fall in the disk centered at (x,y) of radius R. It would also be possible to let R depend on (x,y), i.e., to use different values of R in different subregions of D.

5. Global approximation

As mentioned in the introduction, frequently the data does not warrant constructing an interpolating function (e.g., because of errors). In such cases it may be preferable to construct a surface which only approximates the data. In this section we discuss some global approximation methods.

5.1. Polynomial least squares. The general theory of discrete least-squares fitting is very well known. To briefly review, suppose that $\{\phi_j\}_1^n$ are n given functions on D. Define

$$(5.1)\quad \Phi(a) = \sum_{i=1}^{N} \left| \sum_{j=1}^{n} a_j \phi_j(x_i,y_i) - F_i \right|^2,$$

where $a = (a_1,\ldots,a_n)^T$ is any vector in R^n. Then the problem is to find a^* such that

$$(5.2)\quad \Phi(a^*) = \min_a \Phi(a).$$

The corresponding function

$$(5.3)\quad f(x,y) = \sum_{j=1}^{n} a_j^* \phi_j(x,y)$$

is called the discrete least-squares approximation of the data $\{F_i\}_1^N$. Usually one takes n considerably smaller than N. In this section we briefly discuss least squares using polynomials. Before doing so, however, we make a few general remarks about

solving the general least-squares problem.

There are several approaches to solving (5.2). Perhaps the neatest is the case where the $\{\phi_j\}_1^n$ are orthonormal with respect to the inner-product

$$(5.4) \quad (\phi,\psi) = \sum_{i=1}^{N} \phi(x_i,y_i)\psi(x_i,y_i).$$

Then the solution of (5.2) can be written down explicitly as

$$(5.5) \quad f(x,y) = \sum_{j=1}^{n} F_j\phi_j(x,y).$$

A second very well-known approach to solving (5.2) is via the normal equations

$$(5.6) \quad A^*A\,a = A^*F,$$

where $F = (F_1,\ldots,F_N)^T$ is the vector of data values, and where

$$(5.7) \quad A = (\phi_j(x_i,y_i))_{j=1,i=1}^{n\;,\;N}.$$

In some cases the normal equations are a perfectly acceptable way to compute least-squares approximation, but in other cases the system (5.6) may be ill-conditioned (or even singular--cf. the following subsection for spline least squares). This approach is also not convenient should side conditions be desired (e.g., by imposing actual interpolation at some of the values). For more on the normal equations, see any book on Numerical Analysis.

A more modern method of solving least-squares problems is to use general matrix methods. Specifically, consider the observation equations

$$(5.8) \quad A\,a = F.$$

It can be shown that by applying a series of matrix transformations to this system, one can obtain a set of equations giving the vector a^*. For a complete description of methods of this

type see Lawson and Hanson [129] or Stewart [187]. Matrix methods are quite amenable to the adding of side conditions and can also be designed to take account of rank-deficiency of the matrix A (which corresponds to the case of singular normal equations).

Polynomial discrete least-squares fitting has been widely used for fitting surfaces to data, both scattered and regular. Several authors have developed algorithms for polynomial discrete least-squares fitting of scattered data by constructing orthonormal polynomials (e.g. by Gram-Schmidt orthonormalization). See, for example, Cadwell and Williams [42], Crain and Bhattacharyya [61], and Whitten [201,202]. The latter contains a FORTRAN program.

When the data are more regularly distributed, polynomial least-squares fitting can often be simplified. For example, if the data lie on a grid as in (3.4)-(3.6), then the desired orthogonal polynomials are simply products of the one-dimensional orthogonal polynomials associated with the one-dimensional inner products corresponding to $\{x_i\}_0^{k+1}$ and $\{y_j\}_0^{\ell+1}$ respectively; e.g., see Cadwell [41] or Clenshaw and Hayes [56], as well as the survey papers of Hayes [105,108,109].

There are also special methods for handling data which are not on a grid but instead lie on parallel straight lines. For example, Clenshaw and Hayes [56] have developed methods using expansions in terms of Tchebycheff polynomials (although the method actually only produces an approximation to the least-squares fit rather than the actual minimum).

Polynomial least squares can also be interpreted as multidimensional regression as practiced by statisticians, e.g., see Effroymson [75]. For example, if we are trying to fit a function in the form

$$f(x,y) = \sum_{i=0}^{dx} \sum_{j=0}^{dy} a_{ij} x^i y^j,$$

then by defining new variables by

$$z_{\nu(dy+1)+\mu} = x^{\nu} y^{\mu}, \qquad \begin{array}{l} \nu = 0,1,\ldots,dx \\ \mu = 0,1,\ldots,dy \end{array}$$

we can write

$$f(x,y) = \sum_{i=0}^{d} b_i z_i \ , \qquad d = dxdy + dx + dy,$$

and the problem becomes one of fitting a linear function in several variables.

We close this section by observing that in some cases it may be desirable to consider weighted least squares. In particular, if we have positive weights $w_i > 0$, $i = 1,2,\ldots,N$, then we may replace Φ in (5.1) by

$$I_w(a) = \sum_{i=1}^{N} w_i \left| \sum_{j=1}^{n} a_j \phi_j(x_i, y_i) - F_i \right|^2 .$$

It is interesting to note that the interpolation formula of Shepard discussed in section 3.3 can be interpreted in terms of weighted least-squares fitting. In particular, fix (x,y) in D, and let $r_i(x,y)$ be the distance from (x,y) to the point (x_i, y_i) as before. Now set $w_i = r_i^{-\mu}$, and consider least-squares approximation by a constant c, using these weights. Then one easily computes that the least-squares choice of c is

$$c = \frac{\sum_1^N w_i F_i}{\sum_1^N w_i} = \frac{\sum_1^N F_i r_i^{-\mu}}{\sum_1^N r_i^{-\mu}} .$$

This approach was adopted by Pelto, Elkins and Boyd [152] (as pointed out to me by Chuck Duris).

5.2. Discrete least-squares fitting by splines. As outlined in the previous subsection, discrete least squares can be carried out with any finite set of functions. It is not surprising that a number of authors have tried using tensor product splines. See, e.g., Halliday, Wall, and Joyner [96], Hayes and Halliday [110], Jordan [119], Hanson, Radbill, and Lawson [97], and Whiten [199]. Hayes and Halliday have developed both ALGOL and FORTRAN programs. It is, on the other hand, perhaps somewhat surprising that least-squares fitting with splines can be somewhat problematical. We briefly discuss the method.

Suppose that $H = [a,b] \times [c,d]$ is a rectangle containing the domain D of interest. Let $\{x_i\}_0^{k+1}$ and $\{y_j\}_0^{\ell+1}$ be partitions of $[a,b]$ and $[c,d]$, respectively, and let $\{N_{ij}\}_{1-m,1-n}^{k+1\ \ \ell+1}$ be the tensor product B-splines discussed in section 3.5. We consider discrete least-squares fitting using these $(k+m)(\ell+n)$ B-splines.

To explain how difficulties can arise with spline least-square fitting, we observe that it is quite easy for the matrix A in the observational equations (5.8) to be rank-deficient. On a trivial level this can happen if for some B-spline N_{ij}, none of the data points lies in its support. This deficiency can, of course, be easily removed by dropping this particular B-spline from the set being used to approximate. But rank deficiency can also occur in more subtle ways because of the local support properties of the functions. This problem can be overcome with properly designed algorithms. See Hayes and Halliday [110] for a careful discussion of spline least-squares fitting. Lawson and Hanson [129] include a general discussion of how to handle rank deficient least-squares problems.

If we operate in terms of the normal equations, then it may well occur that the normal equations are in fact singular. This is again due to the local property of the B-splines com-

bined with the discrete inner-product. Even when it is not singular, the set of normal equations can be ill-conditioned (even though it is a relatively sparse matrix with a kind of repeated band-structure).

Discrete least squares can also be carried out with various finite dimensional linear spaces of blended functions. For an extensive study of such methods, see the dissertation of Doty [71].

5.3. Discrete ℓ_1 and ℓ_∞ approximation. Instead of performing discrete least squares, we may consider the following discrete approximation problem: Given functions $\{\phi_j\}_1^n$ defined on D, we seek a^* so that

$$(5.9)\quad \Phi(a) = \sum_{i=1}^{N} \Big| \sum_{j=1}^{n} a_j \phi_j(x_i, y_i) - F_i \Big|$$

is minimized. Alternatively, we may minimize

$$(5.10)\quad \Phi(a) = \max_{1 \le i \le N} \Big| \sum_{j=1}^{n} a_j \phi_j(x_i, y_i) - F_i \Big|.$$

These are the usual ℓ_1 and ℓ_∞ best approximation problems. Both of these problems can easily be reformulated as linear programming problems for the determinations of the optimal a^* (cf. Rabinowicz [160,161] or Rosen [167]). Reasonable choices for the $\{\phi_j\}$ would be low-degree polynomials if D is small, or possibly spline functions.

Discrete approximation methods of this type have had relatively little exposure in the literature. For some results using tensor product splines in the ℓ_∞ problem, see Rosen. The optimal a^* was obtained there by using the standard simplex method on the associated dual linear programming problem.

The ℓ_∞ problem can also be solved by using Remez-type algorithms. For an algorithm which performs generalized

rational approximation (and thus can also be used for polynomial approximations) see Kaufman and Taylor [120]. Theoretical considerations for Tchebycheff approximation in several variables can be found in Collatz [58] or Weinstein [196], for example.

5.4. Spline smoothing (scattered data). In this section we consider some minimization problems similar to those discussed in section 3.4, but where the class of admissible functions is not required to interpolate and where the functional to be minimized includes a term measuring how close the function comes to fitting the data. To be more specific, suppose X is a linear space of "smooth" functions and that Θ is a functional on X which measures the smoothness of an element in X. Suppose in addition that E is a functional defined on X which measures how well a function fits the data. Then the spline-smoothing problem is the following:

$$(5.11)\quad \text{Find } s \in X \text{ such that } \rho(s) = \inf_{u \in X} \rho(u),$$

where

$$(5.12)\quad \rho(f) = \Theta(f) + E(f).$$

The abstract theory of spline smoothing has been well developed; see, e.g., the book of Laurent [127] and references therein. To illustrate the ideas, we briefly discuss a couple of examples. We suppose as in section 3.4 that X is a semi-Hilbert space and that Θ is a seminorm on X with $N = \{f \in X: \Theta(f) = 0\}$. We also suppose that X is actually a function space defined on a domain D, and that the point evaluators at $\{(x_i, y_i)\}_1^N$ are bounded linear functionals on X. We define

$$(5.13)\quad E(f) = p \sum_{i=1}^{N} [f(x_i, y_i) - F_i]^2,$$

where p is a fixed positive constant. Then it can be shown

(cf. Duchon [72,73]) that the solution of Problem (5.11) is a spline which can be written in the form (3.20), where now the coefficients are determined from the linear system

$$(5.14)\qquad \begin{aligned} &\sum_{i=1}^{N} K((x_j,y_j);(x_i,y_i))a_i + \sum_{i=1}^{d} b_i p_i(x_j,y_j) + a_j/p = F_j, && j = 1,2,\ldots,N,\\ &\sum_{i=1}^{d} a_i p_k(x_i,y_i) = 0, && k = 1,2,\ldots,d. \end{aligned}$$

As in section 3.4, the application of this method depends on constructing a reproducing kernel K. If Θ is chosen as in (3.22), Atteia [10-12] and Thomann [192,193] considered spline smoothing for spaces of smooth functions on the rectangle and on the disc (the latter even contains ALGOL programs). Duchon [72,73] considers similar problems defined on $D = R^2$.

A similar spline-smoothing problem has also been considered by Pivorarova [154], where Θ is taken to be

$$(5.15)\qquad \Theta(f) = \iint [D_x^2 f]^2 + [D_y^2 f]^2.$$

See also Kubik [123].

5.5. Smoothing splines (gridded data). In section 3.5 we considered several minimization problems whose solutions led to interpolating polynomial splines (and generalizations). In conjunction with the development of interpolating splines for gridded data, there was a concurrent development of smoothing splines. For example, instead of minimizing the integral Θ in (3.29) over appropriate smooth interpolating functions, we may minimize instead $\rho(f) = \Theta(f) + pE(f)$, where E is given by

$$(5.16)\qquad E(f) = \sum_{i=0}^{k+1} \sum_{j=0}^{\ell+1} [f(x_i,y_j) - F_{ij}]^2.$$

For results in this direction, see e.g. Nielson [149,150]. For

Θ given by (3.29), the smoothing splines are again polynomial splines. Again, more general linear differential operators and more general linear functionals can be considered.

5.6. Continuous least squares. The method of continuous least squares is not directly suited to fitting surfaces to discrete data, but it can be of use as a second-stage process, so we briefly review it. We suppose now that F is a function defined on D which we wish to approximate, and that $\{\phi_j\}_1^n$ are given functions on D. We define

(5.17) $\langle f,g\rangle = \iint_D f(x,y)g(x,y)\,dxdy, \quad \|f\|^2 = \langle f,f\rangle$

and

(5.18) $\Phi(a) = \|\sum_{j=1}^n a_j\phi_j - F\|^2.$

The problem is to find a^* to minimize $\Phi(a)$. The solution is given by solving the normal equations

(5.19) $Aa = r,$

where

$$A = (\langle\phi_i,\phi_j\rangle)_{i,j=1}^n \quad \text{and} \quad r = [\langle\phi_1,F\rangle,\ldots,\langle\phi_n,F\rangle]^T.$$

For reasonably nice approximating functions it is often possible to compute the normal matrix exactly. In practice, the difficulty lies in evaluating the right-hand sides. Generally a quadrature formula is required for this. One advantage of the method would be that if several data-fitting problems are to be solved using the same set of approximating functions, one can do the work of inverting the normal matrix just once and re-use the result as many times as desired.

Reasonable choices for the approximating functions include polynomials, or better yet, tensor product B-splines as in (3.30). Here the singularity problems do not crop up for the splines because we are integrating instead of summing over

finitely many points. The normal matrix in this case has a kind of repeated band structure. The entries can be computed exactly, e.g., by Gaussian quadrature (cf. de Boor, Lyche and Schumaker [38]). Uniform best approximation by tensor products of splines has also been considered, e.g., see Sommer [182].

6. Local approximation methods

As pointed out at the beginning of section 4, there are many advantages which accrue if one uses local methods rather than global ones. In this section we discuss some local approximation schemes.

6.1. Patch methods. As in the case of interpolation, the simplest approach to obtaining local approximation methods is to partition the domain and to define a surface (patch) on each subdomain separately. In particular, suppose that $D = \cup\{D_i\}_1^d$, where D_i are disjoint subsets of D. Then we may seek f in the form

$$(6.1) \quad f(x,y) = \{f_i(x,y), \ (x,y) \in D_i, \quad i = 1,2,\ldots,d.$$

To construct the patch $f_i(x,y)$, we might use the data available in the subregion D_i. In certain cases, however, it may well occur that no data at all are available in the set D_i. In this case we may choose a somewhat larger set $\tilde{D}_i$ of points "near" D_i, and use the data in $\tilde{D}_i$ to construct f_i. For any given method, it should be possible to make the choice of $\tilde{D}_i$ adaptive so that the size of $\tilde{D}_i$ is kept as small as possible consistent with the amount of data desired for the construction of f_i.

The patch method outlined above can be used with any of the approximation methods discussed in section 5. For example, one might choose to use polynomials (of low order), and to do discrete least-squares approximation. Or, one might use ℓ_1 or ℓ_∞ approximation or some other convenient space (e.g. splines)

instead of polynomials. The main point is to keep the size of each individual patch problem (and thus the size of the corresponding system of equations) small. We may have to solve a lot of systems of equations, but each will be small and fairly well-conditioned.

To illustrate how the adaptive feature might be implemented, suppose that the domain D of interest has been enclosed in a rectangle H and that a partition of H is defined by $H = \cup\{H_{ij}\}_{i=0, j=0}^{k, \ell}$, with $H_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ for some

$$(6.2) \quad a = x_0 < x_1 < \ldots < x_{k+1} = b, \; c = y_0 < y_1 < \ldots < y_{\ell+1} = d.$$

Now suppose that we want to do discrete least-squares fitting using a patch of the form $f_{ij}(x,y) = a + bx + cy$ on H_{ij}. In this case it would be reasonable to require that at least 3 pieces of data should be used to construct f_{ij}. If H_{ij} does not contain 3 pieces of data, we expand H_{ij} to $\tilde{H}_{ij}$ by adding all bordering rectangles. If this does not contain 3 pieces, we again add all bordering rectangles, etc. We then compute the discrete least-squares polynomial using the data in $\tilde{H}_{ij}$, but then we use the resulting function only in H_{ij}. The process may be repeated to define each required patch. This kind of adaptive algorithm is very easy to program.

In using patch methods to get local interpolation methods, we concentrated on methods using data at corners of triangles or rectangles, and by choosing appropriate forms for the patches, it was possible to get the individual patches to match together to give a continuous global surface (or with more sophisticated patches, even $C^1(D)$ or higher). Here, however, where the individual patches are determined by approximation, it is not very likely that the patches will match up, and the global surface will generally not even be continuous. For most applications, this is a serious drawback. However, as we shall see in

section 7, patch approximation methods can still be very useful as first-stage methods.

6.2. Direct local methods. In this section we discuss some local methods in which the approximating surface is constructed directly from the data without solving any systems of equations. It will be convenient to pose a more general problem than previously considered.

Let $\mathcal{F}$ be a linear space of functions defined on D, and suppose that $\{\lambda_i\}_1^N$ are linear functionals defined on $\mathcal{F}$. Let $\{\phi_i\}_1^N$ be a prescribed set of functions defined on D. Then we are interested in approximation schemes of the following form:

$$(6.3)\quad QF(x,y) = \sum_{i=1}^{N} \lambda_i F \phi_i(x,y).$$

We can think of this as a surface-fitting problem where the data are given by $F_i = \lambda_i F$, $i = 1,2,\ldots,N$. Given the data, we can write the approximation down immediately.

We also observe that if the ϕ_i have support on small subsets of D, and if each λ_i also has support on the same set, then the formula (6.3) is local. For example, if we take λ_i to be point evaluation at the point (x_i,y_i) and $\phi_i(x,y)$ to be a function with support in a neighborhood of (x_i,y_i), then the approximation formula simply becomes

$$(6.4)\quad QF(x,y) = \sum_{i=1}^{N} F_i \phi_i(x,y).$$

This is very reminiscent of the Lagrange form of interpolation (cf. (4.3)), but unless the ϕ_i are taken to satisfy (4.2), QF will not in fact be an interpolant. For this reason, formulae of the form (6.4) (or more generally (6.3)) are sometimes referred to as quasi-interpolants. Local quasi-interpolants of the form (6.3) can be constructed simply by defining the

functions $\{\emptyset_i\}_1^N$ with local supports. If each of these is continuous (or smooth), then QF will be also.

Although a host of quasi-interpolants can be constructed as outlined above, considerable care must be exercised in order to get methods which give good accuracy (when the original function F is smooth). As observed earlier, this is directly related to making the method exact for polynomials, i.e., such that $QP = P$ for all P in some class of polynomials.

To construct methods of the form (6.3) which apply to scattered data, it is necessary to construct appropriate $\{\emptyset_i\}_1^N$. While a host of methods can be constructed this way, it is not so easy to choose the $\emptyset_i$ to make the method exact for polynomials (which, as we remarked earlier, is directly related to how well the method will approximate smooth functions F). To get methods which do have a reasonable degree of exactness (and a correspondingly good error bound for smooth functions), it is easier to first choose the $\{\emptyset_i\}_1^N$, and then try to find suitable $\{\lambda_i\}_1^N$. While this generally rules out using point evaluators at scattered data, it is possible to construct methods based on point evaluators at appropriate points, and such methods can be useful as <u>second-stage</u> approximations.

To illustrate these ideas, we consider construction of local spline approximation methods following the general treatment in Lyche and Schumaker [131]. Suppose D is enclosed in a rectangle H, and that H is partitioned into subrectangles by a grid as in (6.2). Suppose that $\{N_{ij}\}_{1-m,1-n}^{k\ \ ,\ \ell}$ are the tensor product B-splines associated with this partition (cf. (3.30)). We are now interested in approximation schemes of the form

$$(6.5)\quad QF(x,y) = \sum_{i=1-m}^{k} \sum_{j=1-n}^{\ell} \lambda_{ij} F N_{ij}(x,y).$$

In particular, we are going to consider the question of

constructing formulae of this type which are exact for the class of polynomials $\mathcal{P}_{v,u}$, with some fixed $1 \leq v \leq m$ and $1 \leq u \leq n$. This problem has a very simple algebraic solution if we decide to construct each λ_{ij} in the form

$$(6.6) \quad \lambda_{ij} = \sum_{\nu=1}^{v} \sum_{\mu=1}^{u} \alpha_{ij\nu\mu} \lambda^{x}_{ij\nu} \lambda^{y}_{ij\mu} ,$$

where the $\{\lambda^{x}_{ij\nu}\}_{\nu=1}^{v}$ and $\{\lambda^{y}_{ij\mu}\}_{\mu=1}^{u}$ are linear functionals which apply to functions of x and y alone, respectively. It can be shown (cf. [131]) that given any $\{\lambda^{x}_{ij\nu}\}$ and $\{\lambda^{y}_{ij\mu}\}$ satisfying mild independence assumptions, there exist coefficients $\{\alpha_{ij\nu\mu}\}$ such that the formula (6.5) will be exact for $\mathcal{P}_{vu}$. In fact, these coefficients can easily be explicitly computed.

To give one example, suppose

$$(6.7) \quad \begin{cases} \xi_i = \dfrac{(x_{i+1}+\dots+x_{i+m-1})}{(m-1)}, & i = 1-m,\dots,k \\[2ex] \eta_j = \dfrac{(y_{j+1}+\dots+y_{j+\ell-1})}{(n-1)}, & j = 1-m,\dots,\ell. \end{cases}$$

Then we obtain

$$(6.8) \quad QF(x,y) = \sum_{i=1-m}^{k} \sum_{j=1-n}^{\ell} F(\xi_i,\eta_j) N_{ij}(x,y),$$

a formula which exactly reproduces the bilinear polynomials $\mathcal{P}_{1,1}$. This is the multidimensional (tensor product) version of the Variation Diminishing method of Marsden and Schoenberg; it was studied in some detail in Munteanu and Schumaker [145]. This formula is closely related to the Bezier-type surfaces constructed in Riesenfeld [163] (see also Gordon and Riesenfeld [89]).

We should observe that the way formula (6.5) now stands, it may involve information on F which is taken from data outside of the domain D. This situation can be rectified as follows:

Let

$$(6.9)\quad \Omega = \{(i,j):\ \text{support } \lambda_{ij} \cap D \text{ not empty}\}.$$

Then it can be shown [131] that the method

$$(6.10)\quad QF(x,y) = \sum\sum_{(i,j)\in\Omega} \lambda_{ij} FN_{ij}(x,y)$$

remains exact as long as all functions are restricted to D.

To get higher-order methods, depending only on point evaluations, we proceed as follows. Choose

$$(6.11)\quad \begin{array}{ll} x_i < \tau^x_{ij\nu} < x_{i+m}, & \nu = 1,2,\ldots,v \\ y_j < \tau^y_{ij\mu} < y_{j+n}, & \mu = 1,2,\ldots,u, \end{array}$$

for $i = 1-m,\ldots,k$ and $j = 1-n,\ldots,\ell$. Then if we take $\lambda^x_{ij\nu}$ to be point evaluation at $\tau^x_{ij\nu}$ and $\lambda^y_{ij\mu}$ to be point evaluation at $\tau^y_{ij\mu}$, the coefficients in (6.6) are easily computed. Hints on where the τ's should be placed within the support of the B-splines are given by the error analysis in [131].

We close this section with some historical remarks on the development of local approximation schemes in two dimensions. Early papers include Babuska [14], de Boor and Fix [37], and Fix and Strang [80]. For some methods involving triangular partitions, see Fredrickson [82]. Quasi-interpolants were constructed in de Boor and Fix [37] using point evaluation data, but including derivatives. We have followed Lyche and Schumaker [131] where general linear functionals are considered, and where in particular, methods can be constructed using only point evaluation of the function. (Local integrals etc. would also be possible.) The papers [37] and [131] both contain extensive error bound analyses. It is striking that these local spline approximation methods give optimal order error bounds for smooth functions.

7. Two-stage processes

Many of the methods we have discussed in this paper are only applicable when the data are regularly spaced (and in fact, many surface-fitting methods require specification of derivative data as well as function values). Such methods cannot be applied directly to the scattered data-fitting Problem 1.1. On the other hand, some of the most convenient local interpolating and local approximating methods which do work for scattered data produce surfaces which are not globally smooth (or even continuous). Thus, it seems natural to consider the possibility of constructing two-stage processes in which the first stage uses the scattered data to construct an approximation g, while the second stage uses g to generate data for constructing a surface f (with desirable properties, such as smoothness).

Since it is quite clear how various methods discussed in the earlier sections might be put together to yield two-stage processes, it will suffice to mention just a couple of examples here.

7.1. Interpolation/interpolation. Suppose that we want to construct a piecewise bicubic surface based on data given on a rectangular grid as in (3.4)-(3.6). In each subrectangle H_{ij} the 16 coefficients of the bicubic f (cf. (4.6)) would be determined by the values of f, f_x, f_y, and f_{xy} at each of the four corners. Now since our original data-fitting problem only specifies the values of the function at the grid points, local interpolation cannot be carried out directly. However, we can use the data to provide estimates for the values of f_x, f_y, and f_{xy} at the grid points (i.e., we construct g interpolating the data); then we can use local bicubic interpolation as a second stage. The reader will have no difficulty in imagining ways to produce estimates for these quantities. For some methods which appear in the literature, see the papers of Akima

[5,6], Koelling and Whitten [121], and Späth [183].

7.2. Approximation/interpolation. Instead of making the first-stage process interpolation as in section 7.1, it would also be possible to use an approximating process. For example, one might use least-squares polynomial approximation to construct a patch surface and then use some convenient interpolation process as a second stage. For an example of this type, see Hessing *et al* [114].

7.3. Approximation/approximation. This combination is particularly convenient if we are not concerned about getting an interpolating function. Both stages can be made local. To give an example, recently I have constructed an algorithm for fitting surfaces to scattered data in which the first stage consists of polynomial least-squares patch approximation (with adaptive choice of data--see section 6), and where the second stage consists of direct local tensor product spline approximation. Both stages are local, and the final surface is a tensor product spline. Since the second stage is a direct method, it is very cheap to apply. Experiments with real-life data (e.g. from heart potentials, potential fields, and geological maps--see section 2) have produced very promising results. Details, including an analysis of error bounds, will appear elsewhere. I have also tried alternate versions where the patches are constructed as low-order polynomials which are best approximations in the ℓ_1 or ℓ_∞ sense (via linear programming) again with adaptive choice of data. The results were very similar. Finally, I have also experimented with computing patch approximations, followed by continuous least-squares tensor-product spline approximation. Again, the experiments were promising.

8. Contouring

As indicated in the introduction, frequently the goal in

fitting a surface f to data is to construct a contour map which approximates the contour map of the unknown surface F which produced the data. In this section we discuss some methods for constructing contour maps of a surface f.

8.1. Piecewise linear functions on triangles. When the function f to be contoured is a piecewise linear function defined on triangles (and globally continuous), locating contours reduces essentially to a matter of good bookkeeping. Indeed, if H is the height of the contour of interest, then it is easily seen that for a given triangle T with vertices, A, B, and C,

(8.1) the contour does not pass through T if $H < \min(f(A), f(B), f(C))$ or if $H > \max(f(A), f(B), f(C))$

and

(8.2) the contour intersects exactly two sides of T otherwise.

If case (8.2) holds, it is easy to determine which two sides are intersected and, moreover, by using inverse linear interpolation between vertex values, the points on these sides where the contour crosses can be determined. Specifically, if, for example,

$$f(A) < H < f(B),$$

then the contour crosses the line from A to B at the point on the line which is a distance of

$$\frac{(H-f(A))}{(f(B)-f(A))} \; |B - A|$$

from A. Given the points on two sides of a triangle where the contour line crosses, we can now draw the contour line since it is simply a straight line between the points. An algorithm to carry out this procedure requires enumerating the triangles and vertices and some kind of effective search procedure. There are several available in the literature. For ALGOL programs,

see Heap [111,112]. (An earlier paper of Heap and Pink [113] contains a similar FORTRAN program but only for regular triangularizations.) Lawson [128] discusses a similar algorithm. The algorithms mentioned include two possible approaches: (1) one may start with a triangle where it is known the contour intersects, and trace this contour as far as it goes, or (2) one may simply draw the contour lines in all triangles which have them.

8.2. Piecewise bilinear functions on rectangles. Suppose now that the function f to be contoured is a piecewise (continuous) function on a rectangle partitioned into subrectangles by a grid. Since f is linear in x or y on the edges, it follows that we can again determine whether a contour line of height H crosses an edge by inverse linear interpolation. There is in this case, however, a serious difficulty which does not arise in the case of triangles. It may happen that the height H lies on three or even four sides of the rectangle. In this case, it is possible that two different contour lines pass through the rectangle, and it is not clear how to interconnect the points (see Figure 6).

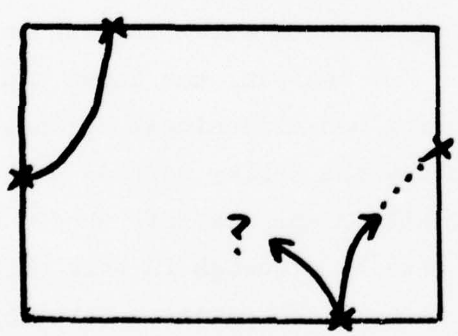


Figure 6. Two Contours in a Rectangle

Put another way, if we are following a contour and enter a rectangle as shown above in Figure 6 on the bottom line, then it is not clear whether we should now turn right or turn left. One approach to designing an algorithm in this case is to simply

always go right, say, even though this may in the end be wrong. (If it is, we have to start over with a coarser mesh.) This technique was incorporated in an algorithm by Heap [111,112]-- the paper contains a FORTRAN program. (An earlier ALGOL program can be found in Heap and Pink [113].

A second approach to overcoming the ambiguity is to compute an approximation to the value of f at the center of the rectangle (e.g., by taking the average of the four-corner values) and then to triangulate the rectangle. This amounts to a second-stage approximation process, and the surface contoured is no longer f itself but an approximation g. This idea was programmed in ALGOL in Heap and Pink [113] and in FORTRAN in Heap [111,112].

Once the set of points for a particular contour have been found, there are a variety of ways of drawing a contour line through these points. One possibility is to simply draw straight lines between each of the points. The actual contour lines are expressions of the form $y = (a+bx)/(c+dx)$ in each rectangle. These are generally not straight lines. Hence, if smoother contours are desired, one may use any one of a number of methods for drawing a smooth curve through an ordered set of points in the plane. For example, the curve could be computed in parametric form using one-dimensional splines. Another possibility would be to use the Bezier methods with either Bernstein polynomials or with B-splines (cf. Gordon and Riesenfeld [89] and Riesenfeld [163]), although in this case the curves will not exactly go through the points. For other algorithms see Marlow and Powell [138] or McConalogue [139].

8.3. Piecewise quadratics on triangles. Suppose now that f is a piecewise quadratic defined on a triangular partition. In this case a contour line at height H passing through a triangle must be described by a conic section. Such a section can

be represented in parametric form as

$$x(t) = (b_0 + b_1 t + b_2 t^2)/(b_3 + b_4 t + b_5 t^2)$$

$$y(t) = (b_6 + b_7 t + b_8 t^2)/(b_3 + b_4 t + b_5 t^2),$$

see Powell [156]. Powell has promised an algorithm based on this observation.

We turn now to some methods for handling general functions f on arbitrary domains D.

8.4. <u>A simple line-printer method</u>. The following simple-minded method can produce reasonable-looking contours without excessive computation, and without recourse to a plotter. Suppose H is a rectangle enclosing the domain D, and that we partition H as $H = \cup H_{ij}$ with a rectangular grid as in (6.2). Let $HL < HU$ be given real numbers. Finally, suppose that t_{ij} is some point in H_{ij} where f can be evaluated (perhaps one of the corners or the center). Define

$$(8.3)\quad C_{ij} = \begin{cases} 0 & , \text{ if } f(t_{ij}) < HL \\ 9 & , \text{ if } f(t_{ij}) > HU \\ v, & \text{if } HL + (v-1)h < f(t_{ij}) < HL + vh,\ 1 \le v \le 8, \end{cases}$$

for all $i = 0,1,\ldots,k$ and $j = 0,1,\ldots,l$ (where $h = (HU-HL)/8$). The $(k+2)$ by $(\ell+2)$ matrix C contains only integers, and if it is printed out without either horizontal or vertical spacing, we obtain a reasonable-looking contour map of the function. A typical example is included in Figure 7. The method can be refined by using an alpha-numeric array C and more than 10 symbols. It can also be refined by using a printer with appropriate horizontal spacing so that each symbol occupies a square rather than a rectangle (e.g., cf. Buneman [40]).

8.5. <u>Threading on a rectangular grid</u>. As in section 8.4,

```
122222222333344444555556666777778888888888867777
2222222222222333344445555566667777788888888888777
2211111111222233344555556667777788888888888888
2211111111111222334445556667777888888899999999
21111ooooo11112233445556667778888889999999999
2111oooooooo112233444556677788889999999999999
2111ooooooooo11223344556677788899999999999999
211oooooooooo11223344556677888999999999999999
211oooooooooo112233445566778889999999999999
211ooooooooo1112234455566778889999999999999
111oooooooo111223344556667788889999999999999
111oooooo11112233445556677788888999999999999
11111111111122334455566677788888899999999999
1111111111222334445566677778888888899998888888
1111111222233344555666777788888888888888888
2222222223333445556667778888888889999999999998
2222223333344455566777788888889999999999999
222333333444455566777788888899999999999999
2333333444455556667778888889999999999999999
3333334444455556667777888888999999999999999
3333334444555566677778888888999999999999999
2333334444555556667777788888889999999999999
2333334444555566666777778888888889999999999
2233334444555566666777778888888888899999999
22333344444555566666777777888888888899999999
22333334444555555666667777777888888889999999
22233334444455555566666677777788888889999999
2223333344444555555666666777777788888899999999
22233333444444555555666667777777888888899999
222233333444445555556666666777777888888889999
22223333334444445555556666666777777788888888899
```

Figure 7. A Simple Contour Map (Heart Potential)

suppose that D is imbedded in a rectangle H which has been partitioned by a rectangular grid as in (6.2). Assuming that f is continuous, it is still possible to decide which of the grid lines a particular contour of height H crosses by examining the end-points of each such line. Since f is not generally linear along such a line, we cannot determine exactly where the crossing point is by linear inverse interpolation. However, if we are willing to evaluate f a few times along this line, we can estimate the crossing point quite accurately by bisection,

for example. Once a sequence of points on a contour has been determined, we may thread a curve through the points just as in section 8.2.

This method does have one serious drawback, however,-- just as with the method discussed in section 8.2--, if we are tracing a contour it may happen that after entering a triangle there is an ambiguity as to which of two points to use to exit the rectangle. One could opt for an ad hoc rule or try the second-stage approximation described in section 8.2. For an example of how this method works, see Falconer [76] (based on Lodwick and Whittle [130]), where it is applied to a surface constructed by local weighted quadratic polynomial least-squares approximation. Since bisection is involved, one should realize that in drawing contours with this routing the surface f is going to be evaluated a great many times.

8.6. Threading on a triangular grid. An obvious cure for the ambiguity discussed in section 8.5 for threading on a rectangular grid is to use a triangular partition in the first place. Then the bisection method coupled with a threading routine leads immediately to a contouring routine for general surfaces f. Strangely enough, I have not been able to find anywhere where this method has been suggested.

I have made no effort to track down all available papers on contouring. A few which I did find and have not yet mentioned are Cottafawa and le Moli [60], Dayhoff [64], and Pelto et al [152]. There are many others.

In some cases it may be desirable to have a more graphic picture of a surface than a contour map can provide. Recently there has been considerable effort devoted to computer methods for displaying surfaces on a scope or with a plotter. For some examples of output and a discussion of methods, see e.g. the book by Barnhill and Riesenfeld [20] on computer-aided design.

If an actual 3-D picture is desired instead of just a perspective, it is even possible to produce holographs.

References

1. Ahlberg, J. H., E. N. Nilson, and J. L. Walsh, The Theory of Splines and Their Applications, Academic Press, New York, 1967.

2. Ahlberg, J. H., E. N. Nilson, and J. L. Walsh, Extremal, orthogonality, and convergence properties of multidimensional splines, J. Math. Anal. Appl. 11 (1965), 27-48.

3. Ahlin, A. C., A bivariate generalization of Hermite's interpolation formula, Math. Comp. 18 (1964), 264-273.

4. Ahuja, D. V., and S. A. Coons, Interactive graphics in data processing, geometry for construction and display, IBM Systems J. 7 (1968), 188-205.

5. Akima, H., Bivariate interpolation and smooth surface fitting based on local procedures, Comm. ACM. 17 (1974), 26-31.

6. Akima, H., A method of bivariate interpolation and smooth surface fitting based on local procedures, Comm. ACM 17 (1974), 18-20.

7. Anderson, W. L., Application of bicubic spline functions to two-dimensional gridded data, U.S. Geological Survey Rpt. GD-71-022, 1971.

8. Arthur, D. W., Multivariate spline functions, I. Construction, properties, and computation, J. Approximation Theory 12 (1974), 396-411.

9. Arthur, D. W., Multivariate spline functions, II. Best error bounds, J. Approximation Theory 15 (1975), 1-10.

10. Atteia, M., Étude de certains noyaux et théorie des fonctions (spline) en Analysis Numérique, thèse, Grenoble, 1966.

11. Atteia, M., Existence et détermination des fonctions splines a plusieurs variables, C. R. Acad. Sci. Paris 262 (1966), 575-578.

12. Atteia, M., Fonctions (spline) et noyaux reproduisants D' Aronszajn-Bergman, Rev. Francaise Informat, Rechereche Operationnelle 4E Annee, R-3 (1970), 31-43.

13. Aziz, A. K. (ed.), The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, Academic Press, New York, 1972.

14. Babuška, I., Approximation by hill functions, Comment.Math. Univ. Carolinae 11 (1970), no. 4, 787-811.

15. Barnhill, R. E., Smooth interpolation over triangles, in [20], 45-70.

16. Barnhill, R. E., Birkhoff, G., and W. J. Gordon, Smooth interpolation in triangles, J. Approximation Theory 8 (1973), 114-128.

17. Barnhill, R. E., and J. A. Gregory, Compatible smooth interpolation in triangles, J. Approximation Theory 15 (1975), 214-225.

18. Barnhill, R. E., and J. A. Gregory, Blending function interpolation to boundary data on triangles, TR 14, Depart. Math., Brunel Univ., 1972.

19. Barnhill, R. E., and L. Mansfield, Error bounds for smooth interpolation in triangles, J. Approximation Theory 11 (1974), 306-318.

20. Barnhill, R. E., and R. F. Riesenfeld, (eds.), Computer Aided Geometric Design, Academic Press, New York, 1974.

21. Bhattacharyya, B. K., An automatic method of computation and mapping of high resolution aeromagnetic data, Geophysics 36 (1971), 695-716.

22. Bhattacharyya, B. K., Bicubic spline interpolation as a method for treatment of potential field data, Geophysics 34 (1969), 402-423.

23. Bhattacharryya, B. K., and B. Raychaudhuri, Aeromagnetic and geologic interpretation of a section of the Appalachian belt in Canada, Canad. J. Earth Sci. 4 (1967), 1015-1037.

24. Biersack, J. P., and D. Fink, Channeling, blocking, and range measurements using thermal neutron induced reactions, in Atomic Collisions in Solids, Vol. 2, edited by S. Datz, B. R. Appleton, and C. D. Moak, Plenum Publishing Corp., New York, 1973, 737-747.

25. Birkhoff, Garrett, Piecewise bicubic interpolation and approximation in polygons, in Approximation with Special Emphasis on Spline Functions, edited by I. J. Schoenberg, Academic Press, New York, 1969, 185-222.

26. Birkhoff, G., and de Boor, C., Piecewise polynomial interpolation and approximation, in Approximation of Functions, H. L. Garabedian, ed., Elsevier, Amsterdam, 1965, 164-190.

27. Birkhoff, G., and H. Garabedian, Smooth surface interpolation, J. Math. Phys. 39 (1960), 258-268.

28. Birkhoff, G., and L. Mansfield, Compatible triangular finite elements, J. Math. Anal. Appl. 47 (1974), 531-553.

29. Birkhoff, G., M. H. Schultz, and R. S. Varga, Piecewise Hermite interpolation in one and two variables with applications to partial differential equations, Numer. Math. 11 (1968), 232-256.

30. Böhmer, K., G. Meinardus, and W. Schempp (eds.), Spline Functions, Lecture Notes 501, Springer-Verlag, New York, 1976.

31. Bohrenberg, G., and E. Giese, Statistical Methoden und ihre Anwedungen in der Geographie, Studienbuch der Geographie, 1975.

32. de Boor, C., Bicubic spline interpolation, J. Math. and Phys. 41 (1962), 212-218.

33. de Boor, Carl, On calculating with B-splines, J. Approximation Theory 6 (1972), 50-62.

34. de Boor, Carl, Appendix to splines and histograms by I. J. Schoenberg, in Spline Functions and Approximation Theory, A. Meir and A. Sharma, eds., ISNM 21, Birkhauser, 1973, 329-358.

35. de Boor, Carl (ed.), Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press, New York, 1974.

36. de Boor, Carl, Splines as linear combinations of B-splines, a survey, this volume.

37. de Boor, C., and G. J. Fix, Spline approximation by quasi-interpolants, J. Approximation Theory 8 (1973), 19-45.

38. de Boor, C., T. Lyche, and L. Schumaker, On calculating with B-splines, II. Integration, to appear Proc. Oberwolfach, 1975.

39. Bramble, J., and M. Zlamal, Triangular elements in the finite element method, Math. Comp. 24 (1970), 809-820.

40. Buneman, O., Sub-grid resolution of flow and force fields, Report no. 452, Inst. for Plasma Research, Stanford Univ., Stanford, California, 1972.

41. Cadwell, J. H., A least squares surface-fitting program, Comput. J. 3 (1961), 266-269.

42. Cadwell, J. H., and D. E. Williams, Some orthogonal methods of curve and surface fitting, Comput. J. 4 (1961), 260.

43. Call, E. S., and F. F. Judd, Surface fitting by separation, J. Approximation Theory 12 (1974), 283-290.

44. Carlson, R. E., and C. A. Hall, Bicubic spline interpolation and approximation in right triangles, J. Math. of Comp., to appear.

45. Carlson, R. E., and C. A. Hall, Ritz approximations to two-dimensional boundary-value problems, Numer. Math. 18 (1971), 171-181.

46. Carlson, R. E., and C. A. Hall, Bicubic spline interpolation in rectangular polygons, J. Approximation Theory 6 (1972), 366-377.

47. Carlson, R. E., and C. A. Hall, Error bounds for bicubic spline interpolation, J. Approximation Theory 7 (1973), 41-47.

48. Carlson, R. E., and C. A. Hall, On piecewise polynomial interpolation in rectangular polygons, J. Approximation Theory 4 (1971), 37-53.

49. Carlson, R. E., and C. A. Hall, Piecewise polynomial approximation of smooth functions in rectangular polygons, J. Approximation Theory 4 (1971), 37-53.

50. Cavendish, J. C., Automatic triangulation of arbitrary planar domains for the finite element method, Int. J. for Numer. Methods in Engineering 8 (1974), 679-696.

51. Chorley, R. J., ed., Spatial Analysis in Geomorphology, Methuen and Co., London, 1972.

52. Ciarlet, P. G., and P. A. Raviart, General Lagrange and Hermite interpolation in $\mathbb{R}^N$ with applications to finite element methods, Arch. Rational Mech. Anal. 46 (1972), 177-199.

53. Ciarlet, P. G., and P. A. Raviart, Interpolation de Lagrange dans R^n, C. R. Acad. Sci. Paris Ser. A273 (1971), 578-581.

54. Ciarlet, P. G., and P. A. Raviart, Interpolation de Lagrange sur des eléments finis courbes dans $\mathbb{R}^n$, C. R. Acad. Sci. Paris Ser. A274 (1972), 640-643.

55. Ciarlet, P. G., and P. A. Raviart, Interpolation theory over curved elements with applications to finite element

methods, Comp. Meth. Appl. Mech. Eng. 1 (1972), 217-249.

56. Clenshaw, C. W., and J. G. Hayes, Curve and surface fitting, J. Inst. Math. Appl. 1 (1965), 164-183.

57. Coatmelec, M. C., Approximation et interpolation des fonctions differentiables de plusieurs variables, Ann. Scient. EC, Norm, Sup. 3, Serie 8 3 (1966), 271-341.

58. Collatz, L., Einschliessungssatz fur die Minimalabweichung bei der Segmentapproximation, Simposio Internazionale Sulle Applicazione dell Analisi alla Fisica Matematica (Cagliari-Sassari, 1964), 11-21.

59. Coons, S. A., Surface patches and B-spline curves, in [20], 1-16.

60. Cottafava, G., and G. Le Moli, Automatic contour map, Comm. ACM 12 (1969), 386-391.

61. Crain, I. K., and B. K. Bhattacharyya, Treatment of non-equispaced two-dimensional data with a digital computer, Geoexploration 5 (1967), 173-194.

62. Davis, J. C., Statistics and Data Analysis in Geology, John Wiley, New York, 1973.

63. Davis, P. J., Interpolation and Approximation, Blaisdell, New York, 1963.

64. Dayhoff, M. O., A contour map program of X-ray crystallography, Comm. ACM 6 (1963), 620-622.

65. Delvos, Franz-Jurgen, On surface interpolation, J. Approximation Theory 15 (1975), 209-213.

66. Delvos, F. J., and H. W. Kösters, On the variational characterization of bivariate interpolation methods, Math. Z. 145 (1975), 129-137.

67. Delvos, F. J., and G. Malinka, Das Blending-Schema von Spline Systemen, in Spline Functionen, K. Bohmer, G. Meinardus, and W. Schempp, eds., Bibliographisches Institut, Zurich, 1974, 47-58.

68. Delvos, F. J., and W. Schempp, Sard's method and the theory of spline systems, J. Approximation Theory 14 (1975), 230-243.

69. Delvos, F. J., and W. Schempp, An extension of Sard's method, in [30], 80-91.

70. Delvos, F. J., and K. H. Schlosser, Das Tensorproduktschema von Spline Systemen, in Splinefunktionen, K. Bohmer, G. Meinardus, and W. Schempp, eds., Bibliographisches Institut, Zurich, 1974, 59-74.

71. Doty, D. R., Blending function techniques with applications to discrete least squares, dissertation, Michigan State Univ., 1975.

72. Duchon, J., Fonctions-spline du type plaque mince en dimension 2, RPT, 231, Univ. of Grenoble, 1975.

73. Duchon, J., Fonctions-spline a energie invariante par rotation, Rpt. 27, Univ. of Grenoble, 1976.

74. Earnshaw, J. L., and I. M. Yuille, A method of fitting parametric equations for curves and surfaces to sets of points defining them approximately, Computer Aided Design 3 (1971), 19-22.

75. Efroysom, M. A., Multiple regression analysis, in Mathematical Methods for Digital Computers, A. Ralston and H. Wilff, eds., Wiley, New York, 1960, 191-203.

76. Falconer, K. J., A general purpose algorithm for contouring over scattered data points, NPL DNAC Rpt. 6, 1971.

77. Ferguson, J., Multivariable curve interpolation, J. Assoc. Comput. Mach. 11 (1964), 221-228.

78. Fisher, S. D., and J. W. Jerome, Elliptic variabional problems in L^2 and L^∞, Indiana J. Math. 23 (1974), 685-698.

79. Fisher, S. D., and J. W. Jerome, Spline solutions to L^1 extremal problems in one and several variables, J. Approximation Theory 13 (1975), 73-83.

80. Fix, George, and G. Strang, Fourier analysis of the finite element method in Ritz-Galerkin theory, Studies in Appl. Math. 48 (1969), 265-273.

81. Frederickson, Paul O., Triangular spline interpolation, Rpt. 6-70, Lakehead Univ, 1970.

82. Frederickson, P. O., Quasi-interpolation, extrapolation, and approximation on the plane, Conf. Numerical Maths., Winnipeg (1971), 159-167.

83. Göel, J. J., Construction of basic functions for numerical utilization of Ritz's method, Numer. Math. 12 (1968), 435-447.

84. Gordon, W. J., Blending-function methods of bivariate and multivariate interpolation and approximation, SIAM J. Numer. Anal. 8 (1971), 158-177.

85. Gordon, W. J., Spline-blended surface interpolation through curve networks, J. Math. Mech. 18 (1969), 931-952.

86. Gordon, W. J., Distributive lattices and the approximation of multivariate functions, in Approximation with Special

Emphasis on Spline Functions, I. J. Schoenberg, ed., Academic Press, New York, 1969, 223-277.

87. Gordon, W. J., Free-form surface interpolation through curve networks, General Motors Research Report GMR-921, 1969.

88. Gordon, W. J., and C. A. Hall, Transfinite element methods. Blending-function interpolation over arbitrary curved element domains, Numer. Math. 21 (1973), 109-129.

89. Gordon, W. J., and R. F. Riesenfeld, B-spline curves and surfaces, in [20], 95-126.

90. Gordon, W. J., and J. Wixom, A note on Shepard's method of metric interpolation to scattered bivariate and multivariate data, to appear.

91. Grant, F., A problem in the analysis of geophysical data, Geophysics 22 (1957), 309-344.

92. Greville, T. N. E., Note on fitting of functions of several variables, SIAM J. Appl. Math. 9 (1961), 109-115.

93. Guenther, R. B., and E. L. Roetman, Some observations on interpolation in higher dimensions, Math. Comp. 24 (1970), 517-522.

94. Hall, C. A., Bicubic interpolation over triangles, J. Math. Mech. 19 (1969), 1-11.

95. Hall, C. A., Natural cubic and bicubic spline interpolation, SIAM J. Numer. Anal. 10 (1973), 1055-1060.

96. Halliday, J., J. F. Wall, and W. D. Joyner, Report on multivariable curve fitting using fundamental splines, Report No. MSN.167, British Aircraft, 1972.

97. Hanson, R. J., J. R. Radbill, and C. L. Lawson, Write-up for SURF, CONT1, SPLVR1, Jet Propulsion Lab., Pasadena, Calif., 1972.

98. Harbaugh, J. W., and D. F. Merriam, Computer Applications in Stratigraphic Analysis, John Wiley, New York, 1968.

99. Haussman, Werner, Mehrdimensionale Hermite-interpolation in Iterationsverfahren, Numerische Mathematik Approximations Theorie, ISNM Vol. 15, L. Collatz, G. Meinardus, H. Unger, and H. Werner, eds., Birkhauser, Basel, 1970, 147-160.

100. Haussman, Werner, On multivariate spline systems, J. Approximation Theory 11 (1974), 285-305.

101. Haussmann, Werner, Tensorproduktmethoden bei mehrdimensionale Interpolation, Math. Z. 124 (1972), 191-198.

102. Haussmann, Werner, Tensorprodukte und mehrdimensionale Interpolation, Math. Z. 113 (1970), 17-23.

103. Haussmann, W., and H. B. Knopp, Divergenz und Konvergenz zweidimensionaler Lagrange-interpolation, to appear Z. Angew. Math. Mech.

104. Haussmann, Werner, and Heinz Josef Munch, On the construction of multivariate spline systems, in Approximation Theory, G. G. Lorentz, ed., Academic Press, New York, 1973, 373-378.

105. Hayes, J. G., Fitting data in more than one variable, in [106], 84-97.

106. Hayes, J. G., ed., Numerical Approximation to Functions and Data, The Athlone Press, London, 1970.

107. Hayes, J. G., New shapes from bicubic splines, NPLNAC 58, also Proceedings of CAD 74: International Conference in Engineering and Building Design, Fiche 36, Row G, International Publishing Co., London, 1974.

108. Hayes, J. G., Available algorithms for curve and surface fitting, in Software for Numerical Mathematics, D. J. Evans, ed., Academic Press, London, 1974, Chapter 12.

109. Hayes, J. G., Numerical methods for curve and surface fitting, Bull. Inst. Math. Applics. (1974), 144-

110. Hayes, J. G., and J. Halliday, The least squares fitting of cubic spline surfaces to general data sets, J. Inst. Maths. Applics. 14 (1974), 89-103.

111. Heap, B. R., Algorithms for the production of contour maps over an irregular triangular mesh, NPL Rpt. NAC 10, 1972.

112. Heap, B. R., Two Fortran contouring routines, NPL NAC 47, 1974.

113. Heap, B. R., and M. G. Pink, Three contouring algorithms, NPL Math. Rpt. 81, 1969.

114. Hessing, R., H. K. Lee, A. Pierce, and E. N. Powers, Automatic contouring using bicubic functions, Geophysics 37 (1972), 669-674.

115. Holroyd, M. T., and B. K. Bhattacharyya, Automatic contouring of geophysical data using bicubic spline interpolation, Geol. Survey Canada Paper No. 70-55, 1970.

116. Hosaka, M., Theory of curve and surface synthesis and their smooth fitting, Information Processing in Japan 9 (1969), 60-68.

117. IMSL Library, International Mathematics and Statistics Library, Houston, Texas.

118. Jentzsch, G., G. Lange, and O. Rosenbach, Anwendung der Spline-funktionen zur Bearbeitung geopnysikalischer Messreihen, in [20], 99-115.

119. Jordan, T. L., Smoothing and multivariable interpolation with splines, LA 3137, Los Alamos, 1964.

120. Kaufman, E. H., and G. D. Taylor, Uniform rational approximation of functions of several variables, Int. J. for Numer. Methods in Engineering 9 (1975), 297-323.

121. Koelling, M. E. V., and E. H. T. Whitten, Fortran IV program for spline surface interpolation and contour map production, Geocomprograms 9 (1973), 1-12.

122. Krumbein, W. C., Trend surface analysis of contour-type maps with irregular control-point spacing, J. Geophysical Res. 64 (1959), 823-834.

123. Kubik, K., Approximation of measured data by piecewise bicubic polynomial functions, manuscript.

124. Kubik, K., The estimation of the weights of measured quantities within the method of least squares, Bull. Geodesique 95 (1970), 21-40.

125. Kubik, K., B. Kunji, and J. Kure, A computer programme for height block adjustment, ITC Rpt., Delft, 1968.

126. Kunz, K. S., Numerical Analysis, McGraw-Hill, New York, 1957.

127. Laurent, P. J., Approximation et Optimisation, Hermann, Paris, 1972.

128. Lawson, C., Generation of a triangular grid with applications to contour plotting, T.M. 299, JPL, 1972.

129. Lawson, C., and R. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, N.J., 1974.

130. Lodwick, G. D., and J. Whittle, A technique for automatic contouring field survey data, Austral. Comput. J. 2 (1970), 104-109.

131. Lyche, T., and L. Schumaker, Local spline approximation methods, J. Approximation Theory 15 (1975), 294-325.

132. Mangeron, D., Sopra un problema al contorno..., Rend. Accad. Sci. Fis. Mat. Napoli 2 (1932), 28-40.

133. Mansfield, L. E., On the optimal approximation of linear functionals in spaces of bivariate functions, SIAM J. Numer. Anal. 8 (1971), 115-126.

134. Mansfield, L. E., On the variational approach to defining splines on L-shaped regions, J. Approximation Theory 12 (1974), 99-112.

135. Mansfield, L. E., On the variational characterization and convergence of bivariate splines, Numer. Math. 20 (1972), 99-114.

136. Mansfield, L. E., Optimal approximation and error bounds in spaces of bivariate functions, J. Approximation Theory 5 (1972), 77-96.

137. Mansfield, L. E., Interpolation to boundary data in triangles with application to compatible finite elements, this volume.

138. Marlow, S. M., and M. J. D. Powell, A FORTRAN subroutine for drawing a curve through a given sequence of data points, Report No. R 7092, A.E.R.E., Harwell, 1972.

139. McConalogue, D. J., An automatic French-curve procedure for use with an incremental plotter, Comput. J. 14 (1971), 207-209.

140. Merriam, D. F., ed., Computer Applications in the Earth Sciences, Plenum Press, New York, 1969.

141. Mitchell, A. R., Introduction to the mathematics of finite elements, in [198], 37-58.

142. Mitchell, A. R., and G. M. Phillips, Construction of basis functions in the finite element method, Nord, Tidskr, Inform. BIT 12 (1972), 81-89.

143. Munteanu, M. J., Generalized smoothing spline functions for operators, SIAM J. Numer. Anal. 10 (1973), 28-34.

144. Munteanu, M. J., On the construction of multidimensional splines, in Spline Functions and Approximation Theory, A. Meir and A. Sharma, eds., ISNM 21, Birkhauser, 1973, 235-265.

145. Munteanu, M. J., and L. L. Schumaker, Some multidimensional spline approximation methods, J. Approximation Theory 10 (1974), 23-40.

146. Nicolaides, R. A., On the class of finite elements generated by Lagrange interpolation, SIAM J. Numer. Anal. 9 (1972), 435-445.

147. Nicolaides, R. A., On a class of finite elements generated by Lagrange interpolation, II, SIAM. J. Numer. Anal. 10 (1973), 182-189.

148. Nielson, G. M., Bivariate spline functions and the approximation of linear functionals, Numer. Math. 21 (1973), 138-160.

149. Nielson, G. M., Multivariate smoothing and interpolating splines, SIAM J. Numer. Anal. 11 (1974), 435-446.

150. Nielson, G. M., Surface approximation and data smoothing using generalized spline functions, thesis, Univ. of Utah, 1970.

151. Norcliffe, G. B., On the use and limitations of trend surface models, Canadian Geographer 13 (1969), 338-348.

152. Pelto, C., T. Elkins, & H. Boyd, Automatic contouring of irregularly spaced data, Geophysics 33 (1968), 424-430.

153. Pilcher, D. T., Smooth parametric surfaces, in [20], 237-254.

154. Pivovarova, N. B., and T. B. Puknacheva, Smoothing experimental data with local splines, Sem. Num. Methods of Applied Mathematics, Novosibirsk, 1975.

155. Poeppelmeir, C., A Boolean sum interpolation scheme to random data for computer aided geometric design, thesis, Univ. of Utah, 1975.

156. Powell, M. J. D., Piecewise quadratic surface fitting for contour plotting, in Software for Numerical Analysis, D. J. Evans, ed., Academic Press, 1974, 253-272.

157. Prenter, P. M., Splines and Variational Methods, Wiley-Interscience, New York, 1975.

158. Prenter, P. M., Lagrange and Hermite interpolation in Banach spaces, J. Approximation Theory 4 (1971), 419-432.

159. Price, J. F., and R. H. Simonsen, Various methods and computer routines for approximation, curve fitting and interpolation, Boeing Rpt. D1-82-0151, 1962.

160. Rabinowitz, P., Applications of linear programming to numerical analysis, SIAM Rev. 10 (1968), 121-159.

161. Rabinowitz, P., Mathematical programming and approximation, in Approximation Theory, A. Talbot, ed., Academic Press, London, 1970, 217-231.

162. Reilly, W. I., Contouring gravity anomaly maps by digital plotter, New Zealand J. Geology Geophysics 12 (1969), 628-632.

163. Riesenfeld, R., Applications of B-spline approximation to geometric problems of computer-aided design, Ph.D. thesis, Univ. of Utah, 1973.

164. Ritter, K., Two-dimensional splines and their extremal properties, Z. Angew. Math. Mech. 49 (1969), 597-608.

165. Ritter, K., Two-dimensional spline functions and best approximations of linear functionals, J. Approximation Theory 3 (1970), 352-368.

166. Robinson, J. E., H. A. K. Charlesworth, and M. J. Ellis, Structural analysis using spatial filtering in interior plans of South-Central Alberta, Amer. Assoc. Petrol. Geol. Bull. 53 (1969), 2341-2367.

167. Rosen, J. B., Minimum error bounds for multidimensional spline approximation, J. Comput. System Sci. 5 (1971), 430-452.

168. Salzer, H. E., Formulas for bivariate hyperosculatory interpolation, Math. Comp. 25 (1971), 119-133.

169. Salzer, H. E., Note on interpolation for a function of several variables, Bull. Amer. Math. Soc. 51 (1945), 279-280.

170. Salzer, H. E., Note on multivariate interpolation for unequally spaced arguments, with application to double summation, SIAM J. Appl. Math. 5 (1957), 254-262.

171. Sard, A., Approximation based on nonscalar observations, J. Approximation Theory 8 (1973), 315-334.

172. Sard, A., Instances of generalized splines, in Spline-Functionen, K. Bohmer, G. Meinardus, W. Schempp, eds., Bibliographisches Institut, Mannheim, 1974, 215-241.

173. Schaback, R., Konstruktion und algebraische Eigenschaften von M-spline Interpolierenden, Numer. Math. 21 (1973), 166-180.

174. Schaback, R., Kollokation mit mehrdimensionalen Splinefunktionen, in Numerische Behandlung nichtlinearer Integrodifferential und Differentialgleichungen, R.Ansorge, and W. Tornig, eds., Lecture Notes 395, Springer-Verlag, Heidelberg, 1974, 291-300.

175. Schlosser, K. H., Mehrdimensionale Spline-interpolation mit Hilfe der Methods von Sard, in [20], 247-264.

176. Schoenberg, I. J., Splines and histograms, in Spline Functions and Approximation Theory, A. Meir and A. Sharma, eds., ISNM 21, Birkhauser, 1973, 277-327.

177. Schultz, M. H., L-infinity-multivariate approximation theory, SIAM J. Numer. Anal. 6 (1969), 161-183.

178. Schultz, M. H., Multivariate L-spline interpolation, J. Approximation Theory 2 (1969), 127-135.

179. Schultz, M. H., Spline Analysis, Prentice-Hall, Englewood Cliffs, N. J., 1973.

180. Shepard, D., A two-dimensional interpolation function for irregularly spaced data, Proc. 1964 ACM Nat. Conf., 517-524.

181. Shu, H., S. Hori, W. R. Mann, and R. N. Little, The synthesis of sculptured surfaces, in Numerical Control Programming Languages, Proc. First Intern. IFIP/IFAC Proclamation Conf., Rome, 1969, North-Holland, Amsterdam, 1970, 358-375.

182. Sommer, M., Alternanten bei gleichmässiger Approximation mit zweidimensional Splinefunktionen, in [20], 339-370.

183. Späth, H., Algorithmus 10- zweidimensionale glatte interpolation, Computing 4 (1969), 178-182.

184. Späth, H., Two-dimensional exponential splines, Computing 7 (1971), 364-369.

185. Späth, H., Two-dimensional glatte interpolation, Computing 4 (1969), 178-182.

186. Steffensen, J. F., Interpolation, Williams and Wilkins, Baltimore, 1927.

187. Stewart, G. W., Introduction to Matrix Computations, Academic Press, New York, 1973.

188. Strang, G., and J. Fix, An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, N. J., 1973.

189. Thacher, H. C., Jr., Derivation of interpolation formulas in several independent variables, Annal. of the New York Acad. of Sciences 86 (1960), 758-775.

190. Thacher, H. C., Jr., and W. E. Milne, Interpolation in several variables, SIAM J. Appl. Math. 8 (1960), 33-42.

191. Theilheimer, F., and W. Starkweather, The fairing of ship lines on a high-speed computer, Math. Comp. 15 (1961), 338-355.

192. Thomann, Jean, Détermination et construction de fonctions spline à deux variables défines sur un domaine rectangulaire ou circulaire, thesis, Lille, 1970.

193. Thomann, J., Obtention de la fonction spline D interpolation A 2 variables sur un domaine rectangulaire ou circulaire, in Proc. Algol en Analyse Numerique II, Centre National de la Recherche Scientifique, Paris, 1970, 83-94.

194. Wachspress, E. L., A rational basis for function approximation, Part 2, J. Inst. Math. Appl. 11 (1971), 83-104.

195. Wachspress, E. L., A Rational Finite Element Basis, Academic Press, New York, 1975.

196. Weinstein, S. E., Approximation of functions of several variables: product Chebychev approximations, J. Approximation Theory 2 (1969), 433-447.

197. Whaples, G. W., A note on degree N independence, SIAM J. Appl. Math. 6 (1958), 300-301.

198. Whiteman, J. R. (ed.), The Mathematics of Finite Elements, Academic Press, London, 1973.

199. Whiten, W. J., The use of multi-dimensional cubic spline functions for regression and smoothing, Austral. Comput. J. 3 (1971), 81-88.

200. Whitten, E. H. T., Scalar and directional field and analytical data for spatial variability studies, Math. Geol. 6 (1974), 183-198.

201. Whitten, E. H. T., Orthogonal polynomial trend surfaces for irregularly spaced data, Int. Assoc. Math. Geol. J. 2 (1970), 141-152.

202. Whitten, E. H. T., More on irregularly spaced data and orthogonal polynomial trend surfaces, Int. Assoc. Math. Geol. J. 4 (1972), 83.

203. Whitten, E. H. T., The practical use of trend-surface analyses in the geological sciences, in Display and Analysis of Spatial Data, Davis and McCullagh, eds., John Wiley, New York, 282-297.

204. Whitten, E. H. T., Orthogonal polynomial contoured trend-surface maps for irregularly spaced data, Tr. 2, Geology Dept., Northwestern Univ., 1974.

205. Whitten, E. H. T., Trends in computer applications in structural geology, in Computer Applications in the Earth Sciences, Plenum Press, New York, 1969, 233-249.

206. Whitten, E. H. T., and M. E. V. Koelling, Computation of Bicubic-spline surfaces for irregularly-spaced data, Tr. 3, Geology Dept., Northwestern Univ., 1974.

207. Whitten, E. T. H., and M. E. V. Koelling, Geological use of multidimensional spline functions, Math. Geology 5 (1973).

208. Whitten, E. H. T., and M. E. V. Koelling, Spline surface interpolation, spatial filtering, and trend surfaces for geologically mapped variables, Math. Geol. 5 (1973), 111-126.

209. Zavialov, Yu.S., An optimal property of bicubic spline functions and the problem of smoothing (Russian), Vycisl, Sistemy, 42 (1970), 109-158.

210. Zavialov, Yu. S., Interpolating L-splines in several variables, Mat. Zametki, 14 (1973), 11-20.

211. Zavialov, Yu. S., Smoothing L-splines in several variables, Mat. Zametki, 15 (1974), 371-379.

212. Zavialov, Yu. S., L-spline functions of several variables, Dokl. Akad. Nauk, SSSR 214 (1974), also Soviet Math. Dokl. 15 (1974), 338-341.

213. Zenisek, A., Interpolation polynomials on the triangle, Numer. Math. 15 (1970), 283-296.

214. Zienkiewicz, O. C., The finite element method: from intuition to generality, Appl.Mech.Rev. 23 (1970), 249-256.

215. Zlamal, M., A finite element procedure of the second order of accuracy, Numer. Math. 14 (1970), 394-402.

216. Zlamal, M., On the finite element method, Numer. Math. 12 (1968), 394-409.

217. Zlamal, M., Curved elements in the finite element method I, II, SIAM J. Numer. Anal. 10 (1973), 229-240; 11 (1974) 347-362.

218. Zwart, Philip B., Multi-variate splines with non-degenerate partitions, SIAM J. Numer. Anal. 10 (1973), 665-673.

Larry L. Schumaker†
Department of Mathematics and Center for Numerical Analysis
The University of Texas at Austin, 78712.

†Supported in part by Grant AFOSR 74-2598B.

APPROXIMATIONS OF ENTIRE FUNCTIONS AND CODING OF SIGNALS WITH A FINITE SPECTRUM

A. G. Vituškin

Functions of real variables whose Fourier transforms have finite support are often called "signals with a finite spectrum." Sound signals, radio signals, areal photographs, holograms and so on are examples of signals of this kind. Coding of a signal usually consists in its discrete representation. In the simplest case, this would be a certain sequence of binary numbers from which the original function can be recovered.

The discrete form of representation rather than the continuous (or the so-called analogue) representation provides more ways of protecting the signal against various kinds of possible distortion in its retaining and transmission. The problem of coding is then reduced to the construction of appropriate approximations. The first requirement which these approximations must satisfy is that the number of binary parameters defining a function should be as small as possible.

It is known that the estimation of the length of the most economical code can be reduced to the calculation of the entropy of the corresponding function class. We will present here the main results concerning the calculation of the entropy of classes of entire functions and discuss one of them in detail.

We will also define a special notion of closeness between signals and present an estimate of the density of codes for the corresponding class of functions.

1 Representation of an entire function and Kotel'nikov's theorem

The concept of signals with finite spectrum is usually associated with the Bernstein class B_σ of entire functions. This class consists, by definition, of the real-valued functions defined and bounded in absolute value by unity on the real line such that the supports of their Fourier transforms are contained in the closed interval $[-\sigma,\sigma]$. Hence, any function f in B_σ is an entire function satisfying the inequality

$$|f(x)| \leq e^{\sigma|\mathrm{Im} z|}$$

for every complex z (cf. [6]). It is known, by Kotel'nikov's theorem [5], that the information content of a signal with spectrum σ is proportional to σ. V. A. Kotel'nikov has also shown in [6] that any square integrable function f in B_σ can be represented in the form

$$f(t) = \frac{\sin \sigma t}{\sigma} \sum_{k=-\infty}^{\infty} (-1)^k f\left(\frac{k\pi}{\sigma}\right) \frac{1}{t - k\pi/\sigma} .$$

It follows from this representation that the number of parameters (per unit time) defining the function is proportional to σ .

Perhaps, the following representation for functions of the class B_σ is more convenient in applications. For any natural number p and $f \in B_\sigma$, we have

$$f(t) = \phi(t)\sin \sigma t + \frac{\sin \sigma t}{\sigma} \sum_{k=-\infty}^{\infty} (-1)^k f\left(\frac{k\pi}{\sigma}\right)$$

$$\left[\frac{1}{t - k\pi/\sigma} - \frac{\sigma}{\pi k} - \left(\frac{\sigma}{\pi k}\right)^2 t - \ldots - \left(\frac{\sigma}{\pi k}\right)^{p+1} t^p\right]$$

where $\phi(t)$ is some polynomial with degree not exceeding p (cf. [6]).

2 The Kolmogorov-Tihomirov theorem

There exist at present more concrete forms of the assertion of Kotel'nikov concerning the information content of signals with a finite spectrum. We consider here two of these results.

DEFINITION 2.1 <u>A set</u> B <u>is called an</u> ε-<u>net of the class</u> B_σ <u>on the interval</u> $[-T,T]$ <u>if for any</u> $f \in B_\sigma$ <u>there exists a function</u> $f^* \in B_\sigma$ <u>such that for any</u> $t \in [-T,T]$

$$|f(t) - f^*(t)| \leq \varepsilon.$$

We denote by $N_\varepsilon(B_\sigma,T)$ the number of elements of the minimal ε-net of the set B_σ on the interval $[-T,T]$ and call

$$H_\varepsilon(B_\sigma,T) = \log_2 N_\varepsilon(B_\sigma,T)$$

the ε-entropy of the set B_σ on the interval $[-T,T]$.

The following theorem due to A. N. Kolmogorov and V. M. Tihomirov can be found in [4].

THEOREM 2.1

$$\underline{\lim}_{\varepsilon \to 0}\ \underline{\lim}_{T \to \infty} \frac{H_\varepsilon(B_\sigma,T)}{2T \log 1/\varepsilon} = \overline{\lim}_{\varepsilon\to 0}\ \overline{\lim}_{T\to\infty} \frac{H_\varepsilon(B_\sigma,T)}{2T \log 1/\varepsilon} = \frac{\sigma}{\pi}.$$

Another model of signals with a finite spectrum is in the form of a random process with a finite spectrum. C. Shannon defined the notion of entropy for a random process and presented (without proof) the asymptotic formulas for the entropy for processes with finite spectrum [7]. It is known that almost all realizations of a process of this kind are entire functions of exponential type. That is why Kolmogorov-Tihomirov's asymptotic formula for the entropy of B_σ has

turned out to be similar to Shannon's. However, Shannon's equality for random processes remained for a long period of time without proof, and it was L. B. Sofman who refined the statement and proved it (cf. [8]).

3 The notion of closeness of signals

We now turn to the main subject of this paper, namely the estimate of the density of codes for signals with a finite spectrum. First we should refine the notion of closeness of signals.

The systems of measurements used in engineering problems are sometimes so complicated that there is no way to describe them by using only the conventional metrics. Let us consider, for example, the system of measurements used in sound recording problems to describe the quality of sound reproduction. Among the technical parameters that are used to characterize the quality of the recording equipment, the following are essential from the point of view of information theory:

1) σ denotes the maximal reproducible frequency,
2) ε is the relative error of the reproduction, and
3) $D = 20 \log_{10}(M/\delta)$ is the dynamic range of the equipment.

Here, M is the norm of the maximal possible output signal and δ is the norm of the noise at the output. The norm of a signal $f(t)$ is defined by

$$\|f(t)\| = \max_t \sqrt{\int_{t-r}^{t+r} f^2(x)\,dx},$$

where $r > 0$ is a constant commensurate with $1/\sigma$.

Taking these parameters as a model, we will give several definitions, the first of which is the definition of closeness.

DEFINITION 3.1. Let ε, δ and r be fixed positive constants and $f(t)$, $f^*(t)$ be functions defined on the whole real line.

We say that the function f^* is close to the function f if for any real t we have

$$|f(t) - f^*(t)| \leq \varepsilon \max_{t-r \, x \, t+r} |f(x)| + \delta.$$

We shall assume that for functions in B_σ the constant r is commensurable to $1/\sigma$. The meaning of this restriction will be clear when we discuss the results below.

4 Complexity of an apparatus

Next, we define the notion of an apparatus and the parameters that are used to characterize its quality and complexity. An apparatus P is a pair of transformations P_1 and P_2 possessing the following properties: Every real-valued function $f(t)$ defined on the whole real axis, called an input function, is transformed by the operator P_1 into a function $\phi = \phi(k\tau, f)$ defined for all integers, k, where τ is a positive constant independent of the input functions; the function ϕ is allowed to take on only the values 0 and 1. In other words, the operator P_1 associates with every input function $f(t)$ a sequence of binary numbers $\phi(k\tau, f)$, $(k = -\infty, \ldots, 0, \ldots, \infty)$, uniformly distributed in time with the density $1/\tau$ per unit time. This sequence is called a binary code of the input function $f(t)$. The second operator P_2 transforms the sequence $\phi(k\tau, f)$ into a real-valued function $f^* = P(f)$ defined on the whole real axis and bounded in modulus by unity.

In addition, it is further assumed that there exists a positive constant ℓ such that for any input function $f(t)$ and every integer k the value $\phi(k\tau, f)$ is uniquely determined by the values of the function $f(t)$ in the interval $[k\tau - \ell, k\tau + \ell]$, and for every t the value $f^*(t)$ depends only on the values $P_1(k\tau)$ for $t - \ell \leq k\tau \leq t + \ell$.

The constant ℓ is called the delay and the number $h = 1/\tau$ the code density of the apparatus. If the boundedness condition of the delay of the apparatus were omitted from the above definition, the notion of code density would not be well-defined. Indeed, by stretching the code sequence we can assign any value to the code density.

The parameters h and ℓ characterize, in a sense, the complexity of the apparatus.

5 Quality of an apparatus

To describe the quality of reproduction we will use three parameters σ, ε and δ.

We say that the parameters of an apparatus P are not worse than σ, ε, δ if for every function $f \in B_\sigma$ the corresponding function $f^* = P(f)$ is close to $f(t)$. In other words, an apparatus has parameters σ, ε and δ if it records and reproduces signals so precisely that for any signal with spectrum σ *the output signal is close to the* input signal.

In engineering problems, for an equipment with parameters σ , ε, δ, the number $D = 20 \log_{10}(1/\delta)$ is called the dynamic range of the equipment. It is said that the equipment has a wide dynamic range, if both large and small signals can be reproduced with the same accuracy.

Having given all the necessary definitions, we can now formulate the result.

6 Estimate of the code density

For any positive numbers σ, ε and δ, it is possible to construct an apparatus with parameters not worse than σ, ε, δ, and with complexity characterized by the inequalities

$$h \leq \frac{\sigma}{\pi} \log \frac{C}{\varepsilon} \text{ and } \ell \leq \max\left(\frac{C}{\varepsilon}, \frac{C}{\delta}\right),$$

where C is an absolute constant. It should be pointed out that the right hand side of the first inequality above does not contain the parameter δ. This means that is is possible to construct an apparatus with an arbitrarily wide dynamic range D by using codes with density independent of D.

This seems rather unexpected, because in engineering problems another point of view prevails: Namely, in constructing an apparatus with the analogue system of recording a sufficiantly wide dynamic range is most difficult to obtain. Hence, we should not expect that a wide dynamic can be obtained without any difficulties at all. In the digital system, for example, obtaining a wide dynamic range requires either long codes or complex schemes of coding. In addition, it should also be noted that it is impossible to construct an apparatus with infinite dynamic range using codes of finite density.

7 Entropy of the class B_σ

The estimate of code density consists, as usual, of counting the entropy of the corresponding function class.

Let the numbers σ, ε, δ and r introduced above be fixed, and let B^* be a set of functions defined on a closed interval $[-T,T]$. This set is called a **net** of the class B_σ on the segment $[-T,T]$, if for any function $f \in B_\sigma$ there exists a function $f^* \in B^*$ close to f on $[-T,T]$; meaning that for any $t \in [-T,T]$ the following inequality holds:

$$|f(t) - f^*(t)| \leq \max_{t-r\leq x\leq t+r} |f(t)| + \delta.$$

Denote by $N(T)$ the number of elements of the minimal net of the set B_σ on $[-T,T]$. The number $H(T) = \log N(T)$ is called the (ε,δ)-entropy of the set B_σ on $[-T,T]$. The following theorem can be found in [3].

THEOREM 7.1 *Let* σ, $\varepsilon \leq 1$, $\delta \leq 1$ *and* $r \geq 1/\sigma$ *be positive numbers*. *Then*

$$H(T) = \frac{2\sigma T}{\pi} \log \frac{C}{\max\{\varepsilon, \delta\}}$$

for all sufficiently large T, *where* C *is a positive function of* σ, ε, δ, r *and lies between two absolute positive constants* C_1 *and* C_2.

Denote by $H = H(\sigma, \varepsilon, \delta)$ the minimum of the code density $h = h(P)$ taken over all apparatuses with parameters σ, ε and δ. It can easily be shown that

$$H = \lim_{T \to \infty} \frac{1}{2T} H(T).$$

Indeed, for any T, on the one hand, any apparatus with parameters σ, ε and δ generates a net of the class B_σ on the segment $[-T,T]$ (this net is the set of all output signals when the input signals are all functions from the class B_σ), and on the other hand, any net can be looked upon as an apparatus which puts in correspondence to every function from B_σ one of the nearest elements of the net.

Hence, the theorem just formulated implies that

$$H = \frac{\sigma}{\pi} \log \frac{C}{\max\{\varepsilon, \delta\}}.$$

That is, the code density of the most economical apparatus with parameters σ, ε, δ is equal to

$$\frac{\sigma}{\pi} \log \frac{C}{\max\{\varepsilon, \delta\}}.$$

8 Discussion of the result

If we put $\delta = 0$ and take ε sufficiently small, then

the constants $1/2^k$ (k running over all positive integers) are pairwise distant, i.e., none of these constants is close to another. Hence, the entropy $H(T)$ is infinite and consequently, $H = \infty$. This means that there is no apparatus with an infinite dynamic range.

Recall that the definition of closeness of signals contains the parameter r. We have been assuming all along that $r \geq 1/\sigma$. If we put $r = 0$, then the corresponding value H turns out to be equal to

$$\frac{\sigma}{\pi} \log \frac{C}{\min\{\varepsilon, \delta\}} ,$$

where the constant C is again understood to be a positive function of all parameters separated from zero and infinity. Hence, in the estimate of H the symbol $\min\{\varepsilon,\delta\}$ replaces $\max\{\varepsilon,\delta\}$. That is, in the case when $r = 0$ and $\delta < \varepsilon$, the code density H of the most economical apparatus turns out to be equal to $(\sigma/\pi)\log(C/\delta)$. We remark that in this case, H happens to be essentially dependent on the parameter δ .

This circumstance shows that the conclusion, that there exists an apparatus with a wide dynamic range and relatively small code density, is correct as long as the choice of metric is reasonable.

The notion of closeness of signals has been defined to correspond to the system of measurements which at present is used in radioengineering. The condition $r \geq 1/\sigma$ seems to be a natural one as well, because errors of reproduction are usually related to the energy of the signal for each certain period of time and not to the momentary value of the signal. For sinusoidal signals, for example, the error is usually related to the energy per period of the oscillation. So there is hope that our choice of metric is resonable

and our conclusion is correct.

9 Estimate of capacity of a communication channel

Our talk about coding has been centered around the sound recording problems. But the estimate presented actually relates to arbitrary signals with finite spectrum and therefore can be used in other applications. For example, the result may be looked upon as the estimate of the capacity of a communication channel.

Any radio communication channel uses signals with finite spectrum and hence can be interpreted as an apparatus. In this case we may use the parameters σ, ε and δ to characterize the frequence range of the chennel, nonlinear distortions of the chennel and the level of channel noise. The entropy $H(T)$ of the corresponding class B_σ characterizes the information content of the signals and the number $H(\sigma, \varepsilon, \delta)$ turns out to be equal to the channel capacity.

The fact that H does not essentially depend on the parameter δ when δ is sufficiently small with respect to ε means that the channel capacity does not depend on the level of channel noise as long as the noise is sufficiently small with respect to the distortions.

10 Estimate of derivatives of polynomials

We now present a result obtained while proving the above theorem. It seems to be of some interest in itself.

Let $P(t)$ be a polynomial of degree k and $M = \max\{|P(t)|: -1 \leq t \leq 1\}$. By Bernstein's theorem the derivative of P at the origin satisfies the inequality $|P'(0)| \leq Mk$. It is well known that this estimate is sharp. V. I. Buslaev [1] has found another form of estimating derivatives as follows:

If the polynomial $P(t)$ has real coefficients, then

$$|P'(0)| \leq A\mu M, \text{ where } \mu = 1 + q + \sum_{i=1}^{k-q} \frac{1}{|r_i|^2},$$

A is an absolute constant, q the number of the zeros of the polynomial P located in the unit disk $|t| \leq 1$ and $\{r_i\}$ are the roots of the polynomial located outside the unit disk.

Polynomials which arise as approximations of entire func- have widely scattered zeros. For such polynomials in particular, this estimate turns out to be much more effective than Bernstein's theorem. However, for polynom‘ .s with complex coefficients in general this estimate is not valid. This can be seen from the counter-example

$$P(t) = \left(1 + \frac{it}{\sqrt{k}}\right)^k .$$

In [2] Buslaev presented an estimate of the derivative of a polynomial in terms of the characteristic μ in the integral metrics and a generalization of Nikolski's inequality relating norms of a polynomial in various metrics.

11 Scheme of construction of nets

Let us assume that $\sigma = \pi$ and $r = 1$. The representation of entire functions cited in section 1 asserts the possibility of approximating a function $f \in \mathbf{B}_\pi$ uniformly on $[-T-1, T+1]$ for large T to within δ by a function of the special form $R(t)P(t)$. Here, the factor $R(t)$ is a universal function not depending on the choice of the function being approximated, and the factor $P(t)$, which depends on the choice of the function being approximated, is a polynomial of degree $n \leq 2T + o(T)$. The problem is thus reduced to the calculation of the quantity $H_{\varepsilon,\delta}$ for the class of functions of the form $R(t)P(t)$. The function $R(t)P(t)$

AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1

APPROXIMATION THEORY II

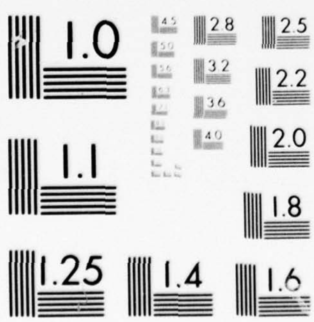
NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.

UNCLASSIFIED AFOSR TR-76-1242 N/L

4 OF 7

AD
A033032

033032



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

is uniquely determined (to within a constant factor) by the collection $A = \{a_1, \ldots, a_n\}$ of the zeros of the polynomial $P(t)$. If two collections of zeros A and $\tilde{A}$ are close in a certain sense, then the function $R(t)\tilde{P}(t)$ determined by the collection $\tilde{A}$ is $(\varepsilon,0)$-close to the function $R(t)P(t)$ determined by the collection A. The accuracy of approximations of zeros is characterized in terms of the density of the zeros. A special characteristic μ is introduced (see section 10) to describe the density of a collection of zeros.

The most essential problem in computation of the entropy is the finding of an upper bound. A lower estimate can easily be obtained from the estimate of the ε entropy of B_σ in the uniform metric (cf. [4]).

As we see from the above scheme, the construction is based on the possibility of the representation of polynomials in one variable by producing linear functions. Polynomials in several variables do not have the same kind of representations. This is the main obstacle for obtaining an asymptotic formula for the (ε,δ)-entropy of the class of entire functions in several variables.

References

1 Buslaev, V. I., An estimation of the derivative of a polynomial with real coefficients, Izv. Akad. Nauk SSSR, Ser. Mat. 39 (1975), 413-417.

2 Buslaev, V. I., Some inequalities for polynomials with real coefficients, Dokl. Akad. Nauk SSSR, 223 (1975) 20-22; English transl., Soviet. Math. 16 (1975), 803-805.

3 Buslaev, V. I., and A. G. Vituškin, An estimate of the length of a signal code with a finite spectrum in connection with sound transcription problems, Izv. Akad. Naul SSSR, Ser. Mat. 38 (1975) 867-895, English transl., Math. USSR Izvestija 8 (1974), 867-894.

4 Kolmogorov, A. N., and V. M. Tihomirov, ε-entropy and ε-capacity of sets in functional spaces, Uspehi Mat. Nauk 14 (1959), no. 2 (86), 3-86; English transl., Amer. Math. Soc. Transl. (2) 17 (1961), 277-364.

5 Kotel'nikov, V. A., Materials for the first All-Union Congress on questions of the reconstruction of the communication system and of the development of weak current industry, Izdat. Upravl. RKKA, 1933.

6 Levin, B. Ja., Entire functions (Course of lectures), Inst. Meh. Moskov, Gos. Univ., Moscow, 1971 (Russian).

7 Shannon, C. E., The mathematical theory ofcommunication, Bell Syst., Techn. Journal, 27 (1948), 379-423.

8 Sofman, L. B., A problem of Shannon, Dokl. Akad. Nauk SSSR, 215 (1974), 1313-1316: English transl., Soviet Math. Dokl. 15 (1974), 704-708.

A. G. Vituškin
Steklov Mathem. Institute
Moscow, USSR.

ON THE CONVERGENCE OF SMOOTH MONOSPLINES TO POLYNOMIAL MONOSPLINES

R. B. Barrar and H. L. Loeb

1 Introduction

In many problems in numerical analysis optimal monosplines play an important role, [1,2,5,8,9]. For example in several applications one wants to characterize the monosplines of least norm where both the knots and coefficients are free parameters but the multiplicities are fixed odd positive integers. We phrase this monospline problem in the following way. Let $K(x,y)$ be a given real-valued function on $[0,1] \times [0,1]$. Further consider the fixed positive integers $\{n, m_1, \ldots, m_t\}$ where m_i is odd and is called a multiplicity. We can state our problem as follows:

Problem I. Among all monosplines of the form

$$(1) \quad M(x) = \int_0^1 K(x,y)dy + \sum_{i=1}^{t} \sum_{j=0}^{m_i-1} a_{ij} K^{(j)}(x, \xi_i) + \sum_{j=0}^{n-1} a_j K^{(j)}(x,0)$$

where the knots satisfy $0 < \xi_i < \xi_{i+1} < 1$ and the coefficients a_j, a_{ij} are real numbers, find the $M(x)$ of minimal L_p norm on $[0,1]$ $(1 \leq p \leq \infty)$. Here

$$K^{(j)}(x, \xi_i) = \frac{\partial^j}{\partial \xi^j} K(x,\xi)\Big|_{\xi=\xi_i}.$$

If we allow the knots in Problem I to coalesce we have:

Preceding page blank

Problem II. Among all monosplines of the form

$$(2)\quad \int_0^1 K(x,y)dy + \sum_{i=1}^{t'} \sum_{j=0}^{m'_i-1} a_{ij}K^{(j)}(x,\xi_i) + \sum_{j=0}^{n-1} a_j K^{(j)}(x,0)$$

where $\{m'_1,\ldots,m'_{t'}\}$ are any set of positive integers so that for some integers $0 = r_1 < r_2 < \ldots < r_{t'+1} = t$ we have

$$m'_i = \sum_{j=r_i+1}^{r_{i+1}} m_j \quad (i = 1, \ldots, t'),$$

we seek the monospline of smallest L_p norm ($1 \leq p \leq \infty$). Recall $(m_1,\ldots,m_t)$ are a fixed set. Note that the class of functions defined in Problem I is a subset of the functions considered in Problem II.

In the first portion of this paper we show that if $K(x,y)$ is a smooth extended totally positive kernel [6, p.49] then Problem II always has a solution. Further we prove that any solution to Problem II is a solution to Problem I and we characterize all these solutions.

In the second portion of the paper we consider the polynomial monospline kernel,

$$K_o(x,y) = n(x - y)_+^{n-1}$$

where

$$x_+^{n-1} = \begin{cases} x^{n-1} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

It is well known that $K_o(x,y)$ is a totally positive kernel, [6, p.511]. We seek a solution to Problem I in the uniform norm when $K_o(x,y)$ is used as the kernel and each multiplicity, m_i, satisfies

$$m_i \leq n - 1.$$

Since $K_o(x,y)$ is not extended totally positive and also not smooth we attack this problem by considering for each $\varepsilon > 0$ the smooth kernel $K_\varepsilon(x,y)$ obtained from $K_o(x,y)$ by employing the Gaussian Transform; that is,

$$K_\varepsilon(x,y) = \frac{1}{\sqrt{2\pi}\,\varepsilon} \int_{-\infty}^{\infty} \exp[-\frac{1}{2\varepsilon^2}(x - \xi)^2] K_o(\xi,y)d\xi.$$

Karlin [6, p.512] has shown that $K_\varepsilon(x,y)$ is analytic and extended totally positive. Our earlier results show then that we have at least one solution for the smooth kernel and each solution, $M_\varepsilon(x)$, has the property that there are alternation points,

$$0 \leq x_1 < x_2 < \ldots < x_d \leq 1$$

where

$$d = n + \sum_{i=1}^{t} (m_i + 1) + 1$$

so that

$$(3) \quad M_\varepsilon(x_i) = -M_\varepsilon(x_{i+1}) \quad i = 1,\ldots,d-1$$

$$\|M_\varepsilon\| = |M_\varepsilon(x_1)|$$

and we can demonstrate that in this setting $x_1 = 0$ and $x_d = 1$. Our principal result is that for some sequence $\{\varepsilon_i\}\downarrow 0$, $M_{\varepsilon_i}(x)$ converges uniformly to an optimal polynomial monospline $M_o(x)$. Since $M_o(x) \not\equiv 0$, it follows that $M_o(x)$ also satisfies (3), nontrivially.

2 Main results

We will state our main theorems without detailed proofs. The proofs will appear elsewhere.

THEOREM 1. *Let* $K(x,y)$ *be a smooth extended totally positive kernel. Then for each* $1 \leq p \leq \infty$; *Problem II has a solution. Further, each solution is of the form* (1) *where*

$$(4) \quad a_{i,m_{i-1}} < 0 \quad (i = 1,\ldots,t).$$

In the uniform case each best approximation satisfies (3).

Proof. Using the techniques of [3] one can show that a best approximation exists for Problem II. In fact, the only new technique needed to complete the proof involves a parametric method for separating a set of knots which have coalesced. The concept of "extended varisolvence" [4] and Newton's Identity turn out to be quite helpful in this instance.

For each $\varepsilon > 0$, let $M_\varepsilon(x)$ be a best monospline approximation to Problem I in the uniform norm for the kernel $K_\varepsilon(x,y)$. Further let $M_\varepsilon^{(0)}(x)$ be the polynomial monosplines which is obtained from $M_\varepsilon(x)$ by replacing $K_\varepsilon(x,y)$ by $K_o(x,y)$. By going to a subsequence we can assume as $\varepsilon \to 0$, the knots associated with each $M_\varepsilon^{(0)}(x)$ converge. We consider the form,

$$M_\varepsilon^{(0)}(x) = x^n + \sum_{i=0}^{n-1} a_{i,\varepsilon} x^i + S_{n-1,\varepsilon}(x) + S_{n,\varepsilon}(x) + S_\varepsilon(x)$$

where we have developed $M_\varepsilon^{(0)}(x)$ in terms of divided differences. Here $S_{n-1,\varepsilon}(x)$ involves differences up to order $n - 1$ of all knots which have limit points in $(0,1)$. $S_{n,\varepsilon}(x)$ consists of the B splines [6, p.529] for these "interior knots." $S_\varepsilon(x)$ involves all the terms which have knots which converge to "0" or "1".

THEOREM 2. For $0 < \varepsilon \leq \varepsilon_o$, the sequence,

$$\{x^n + \sum_{i=0}^{n-1} a_{i,\varepsilon} x^i + S_{n-1,\varepsilon}(x)\},$$

is uniformly bounded.

Proof. The fact that the Gaussian Transform is variation diminishing [6, p.20,103] plays a key role in the proof.

THEOREM 3. For some sequence $\{\varepsilon_i\}\downarrow 0$ $M^{(0)}_{\varepsilon_i}(x)$ converges uniformly to a $M_o(x)$ of the form (1). Further $M_o(x)$ satisfies (3) and (4) and is a solution to Problem I in the uniform norm for the kernel $K_o(x,y)$.

Proof. A key ingredient in the proof is a sharp upper bound on the number of alternations $M_o(x)$ loses because of the fact that the knots coalesce. The theory of B splines [6, p.531] and Theorem 2 are used to establish this bound.

This last theorem partially answers a question posed by Micchelli, [8]. We note that if all the $m_i = 1$, $M_o(x)$ is the unique minimizer, [10], and we find in this situation that any sequence $\{\varepsilon_i\}\downarrow 0$ has the property $M_{\varepsilon_i}(x)$ converges uniformly to M_o (not just a subsequence).

References

1 Barrar, R. B. and H. L. Loeb, Multiple zeros and applications to optimal linear functionals, to appear Numerische Mathematik.

2 Barrar, R. B., H. L. Loeb and H. Werner, On the existence of optimal integration formulas for analytic functions, Numer. Math., 23 (1974), 105-117.

3 Barrar, R. B. and H. L. Loeb, On a non-linear characterization problem for monosplines, to appear J. Approximation Theory.

4 Barrar, R. B. and H. L. Loeb, On extended varisolvent families, J. D'Analyse Mathematique 26 (1973), 243-254.

5 Karlin, S., Best quadrature formulas and splines, J. Approximation Theory, 4 (1971), 59-90.

6 Karlin, S., Total Positivity, Stanford University Press, Stanford 1968.

7 Karlin, S. On a class of best nonlinear approximation problems, Bull. Amer. Math. Soc. 78 (1972), 43-49.

8 Micchelli, C. The fundamental theorem of algebra for monosplines with multiplicities, Linear Operators and Approximation, edited by P. L. Butzer et al., Birkhauser, Verlag, Basil (1972), 419-430.

9 Schoenberg, I. J., Monosplines and quadrature formulae, Theory and Applications of Spline Functions, edited by T. N. E. Greville, Academic Press, New York (1969), 157-207.

10 Johnson, R. S., On monospline of least deviation, Trans. Amer. Math. Soc. 96 (1960), 458-477.

11 Schumaker, Uniform approximation by Chebyshev spline functions, II; free knots, SIAM J. Numer. Anal. 5 (1968), 647-656.

R. B. Barrar
Department of Mathematics
University of Oregon
Eugene, Oregon 97403

H. L. Loeb
Department of Mathematics
University of Oregon
Eugene, Oregon 97403

QUANTITATIVE KOROVKIN THEOREMS FOR L_p-SPACES

H. Berens and R. A. DeVore

It is the aim of the paper to formulate and prove a quantitative Korovkin theorem for sequences of positive linear contractions on the space $L_p[a,b]$, $1 \leq p < \infty$, and to examine in what sense the given estimates are best possible.

1 Introduction

Korovkin's famous theorem on positive linear operators can be put into a quantitative form.

Let $C[a,b]$ be the space of all continuous real-valued functions on the compact interval $[a,b]$ endowed with the supremum norm $\|\cdot\|_\infty$, and let $(L_n: n = 1,2\ldots)$ be a sequence of positive linear contractions on $C[a,b]$. One version of Korovkin's theorem is the following:

Let $\phi_i(x) = x^i$, $i = 0,1,2$, and

$$\lambda_n^2 = \max_{i=0,1,2} \|L_n\phi_i - \phi_i\|_\infty.$$

For any $f \in C[a,b]$

$$(1) \quad \| L_n f - f \|_\infty \leq C\, \{\lambda_n^2 \| f \|_\infty + \omega_2(f;\lambda_n)\}, \quad n = 1,2,\ldots,$$

with $\omega_2(f)$ the second order modulus of smoothness of f and C a constant depending only on the interval.

If, in particular, $L_n\phi_i = \phi_i$, $i = 0,1$, for $n = 1,2,\ldots$, then $\lambda_n^2(x) = L_n\phi_2(x) - \phi_2(x) \geq 0$ and the following pointwise estimate holds true

$$| L_n f(x) - f(x) | \leq C\, \omega_2(f;\lambda_n(x)),$$

with C again only dependent on the interval.

For translation invariant operators on $C_{2\pi}$ Korovkin himself proved a quantitative theorem, so did T. Popoviciu for Bernstein-type operators already in 1950. For arbitrary sequences of positive linear operators R. G. Mamedov and, independently but later, O. Shisha and B. Mond gave estimates involving the modulus of continuity of the function. The theorem stated above is due to G. Freud [5] in 1968. For further references see [3].

It was observed almost coincidentally by Ralf L. James, by D. E. Wulbert and W. Kitto, and by G. G. Lorentz and one of the authors [1] that the set $\{\phi_o,\phi_1\}$ already forms a "test set" for sequences of positive linear contractions on the Lebesgue spaces $L_p[a,b]$, $1 \le p < \infty$. In view of the quantitative theorem stated above, it is natural to ask whether there are analogue estimates for the L_p-spaces involving the modulus of smoothness and the order of approximation of the functions ϕ_i, $i = 0,1$.

It is the aim of the paper to formulate and prove such theorems and to examine in what sense the given estimates are best possible.

2 Main Results

For an $f \in L_p[a,b]$, $1 \le p < \infty$, we denote its modulus of continuity by $\omega_{1,p}(f)$, and its second order modulus of smoothness by $\omega_{2,p}(f)$; the latter being defined by

$$\omega_{2,p}(f;\delta) = \sup_{0<t\le\delta} \{\int_{S_t} |f(x+t)-2f(x)+f(x-t)|^p dx\}^{1/p},$$

where $S_t = [a+t,b-t]$, $0 < t \le (b-a)/2$.

Let $(L_n: n = 1,2,\ldots)$ be a sequence of positive linear contractions on $L_p[a,b]$, and let

(2) $\lambda_{n,p} = \{\max_{i=0,1} \|L_n\phi_i - \phi_i\|_p\}^{1/2}$.

THEOREM. *For any* $f \in L_p[a,b]$ *and* $n = 1,2,\ldots$

(3) $\|L_n f - f\|_p \le C\{\lambda_{n,p}^{2/p}\|f\|_p + \omega_{2,p}(f;\lambda_{n,p}^{1/p})\}$.

If, in addition, $L_n\phi_o \equiv \phi_o$, then

(4) $\|L_n f - f\|_p \le C'\{\lambda_{n,1}^{2/p}\omega_{1,p}(f;b-a) + \omega_{2,p}(f;\lambda_{n,1}^{1/p})\}$.

The constants are only dependent on the interval $[a,b]$ *and on* p.

The proof depends on two observations, formulated in the following two lemmas. The first one is that for L_p-spaces Theorem 11 in [1] can be given a quantitative form.

As usual, we denote by f_+ the positive part of the function f.

LEMMA 1. *For any* $f \in L_p[a,b]$, $1 \le p < \infty$, *and* $n = 1,2,\ldots$

(5) $\|L_n f_+ - f_+\|_p \le p^{1/p}\|L_n f - f\|_p^{1/p}\|f\|_p^{1-1/p} + \|L_n f - f\|_p$.

Proof. Since $\|L_n f_+ - f_+\|_p \le \|L_n f_+ - (L_n f)_+\|_p + \|(L_n f)_+ - f_+\|_p$ and $\|(L_n f)_+ - f_+\|_p \le \|L_n f - f\|_p$, estimate (5) is proved if we can show that

(6) $\|L_n f_+ - (L_n f)_+\|_p^p \le p\|L_n f - f\|_p\|f\|_p^{p-1}$.

Observe first that when $g,h \ge 0$, $\|g\|_p^p + \|h\|_p^p \le \|g + h\|_p^p$. Taking $g = (L_n f)_+$ and $h = L_n f_+ - (L_n f)_+ \ge 0$, it follows that

$$\|L_n f_+ - (L_n f)_+\|_p^p \le \|L_n f_+\|_p^p - \|(L_n f)_+\|_p^p .$$

Since L_n is contractive,

$$\|L_n f_+ - (L_n f)_+\|_p^p \le \|f_+\|_p^p - \|(L_n f)_+\|_p^p$$
$$\le p(\|f_+\|_p - \|(L_n f)_+\|_p)\,\|f_+\|_p^{p-1},$$

from which inequality (6) easily follows.

REMARK. For $p = 1$ inequality (5) simply reads $\|L_n f_+ - f_+\|_1 \le 2\,\|L_n f - f\|_1$. Consequently, if $\{\gamma_n \ge 0,\ n = 1,2,\dots\}$ is any (null) sequence, then the class of functions

$$\{f \in L_1[a,b] : \|L_n f - f\|_1 \le C_f \gamma_n,\ n = 1,2,\dots\}$$

is a vector sublattice in $L_1[a,b]$. This does not hold true for $1 < p < \infty$ as the examples given in Sec. 3 will show.

For $1 \le p \le \infty$, let

$$W_{2,p}[a,b] = \{f \in L_p[a,b] : f(x) = \int_a^b (x-u)_+ g(u)\,du \ \text{ a.e.},$$
$$g \in L_p[a,b]\}.$$

For $f \in W_{2,p}[a,b]$, f and f' are absolutely continuous on $[a,b]$ and the associated function g equals f''. In other words, $W_{2,p}[a,b]$ equals $L_p^{(2)}[a,b]$ modulo the linear functions. Furthermore, let $V_{2,1}[a,b]$ denote the relative completion of $W_{2,1}[a,b]$, i.e.,

$$V_{2,1}[a,b] = \{f \in L_1[a,b] : f(x) = \int_a^b (x-u)_+ d\mu(u) \ \text{ a.e.},$$
$$\mu \text{ a function of bounded variation on } [a,b]\}.$$

Clearly, for such an f, we have $f' = \mu$.

LEMMA 2. *For* $f \in W_{2,p}[a,b]$, $1 \le p < \infty$,

(7) $\quad \|L_n f - f\|_p \le C\lambda_{n,p}^{2/p}\,\|f''\|_p, \quad n = 1,2,\dots$

More generally, for $f \in V_{2,1}[a,b]$

(8) $\quad \|L_n f - f\|_p \le C\lambda_{n,p}^{2/p}\,[\mathrm{Var}\ f']_a^b, \quad n = 1,2,\dots$

The constant C depends only on the space $L_p[a,b]$.

Proof. If $f \in W_{2,p}[a,b]$, it follows from its very definition that

$$\|L_n f - f\|_p \leq \int_a^b \|L_n((\cdot -u)_+;\cdot) - (\cdot -u)_+\|_p |f''(u)|du.$$

Setting

$$(9) \quad \alpha^2_{n,p} = \sup_{a\leq u\leq b} \|L_n((\cdot -u)_+;\cdot)-(\cdot -u)_+\|_p, \quad n = 1,2,\ldots$$

we obtain

$$\|L_n f - f\|_p \leq \alpha^2_{n,p} \|f''\|_1 \leq (b-a)^{1-1/p} \alpha^2_{n,p} \|f''\|_p.$$

By Lemma 1, $\alpha_{n,p} \leq \text{const.}\ \lambda^{1/p}_{n,p}$, this proves (7). The proof of (8) is similar.

REMARK. $V_{2,1}[a,b]$ forms a normed linear space under the norm $[\mathrm{Var}\ f']^b_a$. Observe that the approximation behavior of a sequence (L_n) on the unit ball of $V_{2,1}[a,b]$ is determined by the behavior of the extreme points of the ball. Here the extremal functions are given by $e_u(x) = \pm(x-u)_+$, $a \leq u < b$. Hence

$$\sup\{ \|L_n f - f\|_1 : f \in V_{2,1} \text{ and } [\mathrm{Var}\ f'] \leq 1\}$$
$$= \sup_{a\leq u<b} \|L_n((\cdot -u)_+;\cdot) - (\cdot -u)_+\|_1.$$

By Lemma 1, $\|L_n((\cdot -u)_+;\cdot) - (\cdot -u)_+\|_1 \leq 2 \|L_n((\cdot -u),\cdot) - (\cdot -u)\|_1 \leq 2 (1+\max(|a|,|b|))\lambda^2_{n,1}$. This shows clearly that for functions $f \in V_{2,1}[a,b]$ the approximation behavior of $L_n f$ towards f is determined by that for ϕ_o and ϕ_1.

Introducing on $L^{(2)}_p[a,b]$, $1 \leq p \leq \infty$, the norm $\|f\|_{2,p} = \|f\|_p + \|f''\|_p$, it follows from Lemma 2 that

$$(10) \quad \|L_n f - f\|_p \leq C\alpha^2_{n,p} \|f\|_{2,p} < C'\lambda^{2/p}_{n,p} \|f\|_{2,p}, \quad n = 1,2,\ldots$$

Proof of the theorem. Let $K_p(t;f) := K_p(t;f;L_p L^{(2)}_p)$, $t > 0$, denote the Peetre K-function norm of the function $f \in L_p[a,b]$, i.e.,

$$K_p(t;f) = \inf \{ \|f-g\|_p + t\|g\|_{2,p} : g \in L_p^{(2)}[a,b] \}.$$

By a well-known result on moduli of smoothness, see H. Johnen [7] or [4],

$$K_p(t^2;t) \asymp \min(1,t^2)\|f\|_p + \omega_{2,p}(f;t). \quad †)$$

Hence for any $f \in L_p[a,b]$ and any $g \in L_p^{(2)}[a,b]$

$$\|L_n f - f\|_p \le \|L_n(f-g) - (f-g)\|_p + \|L_n g - g\|_p$$
$$\le 2\|f-g\|_p + C\alpha_{n,p}^2 \|g\|_{2,p}.$$

Taking the infimum over all $g \in L_p^{(2)}[a,b]$,

$$(11) \quad \|L_n f - f\|_p \le \max(2,C)\, K_p(\alpha_{n,p}^2;f)$$
$$\le C'\{\alpha_{n,p}^2 \|f\|_p + \omega_{2,p}(f;\alpha_{n,p})\}.$$

As was remarked earlier, $\alpha_{n,p} \le \text{const.}\ \lambda_{n,p}^{1/p}$. Putting this in our last estimates gives the theorem.

To prove the estimate (4), let us remark that, since $L_n\phi_o \equiv \phi_o$, $|L_n((\cdot-u)_+;x) - (x-u)_+| \le 2\,(b-a)$ a.e., and, consequently, $\alpha_{n,p}^2 \le 2^{2-1/p}(b-a)^{1-1/p}\lambda_{n,1}^{2/p}$, where $\lambda_{n,1}^2 = \|L_n\phi_1 - \phi_1\|_1$. Furthermore, for any constant-valued function, say, c,

$$\|L_n f - f\|_p \le C\{\alpha_{n,p}^2 \inf_c \|f-c\|_p + \omega_{2,p}(f;\alpha_{n,p})\}.$$

Since $\inf\{\|f-c\|_p : c \in R\} \le \text{const.}\ \omega_{1,p}(f;b-a)$ the last estimate implies inequality (4).

REMARK. A more careful analysis of the constants appearing in the proof will show that constants C and C' can be chosen which only depend on the interval $[a,b]$ but not on the index p.

---

†) $f(t) \asymp g(t)$ means $f(t) = O(g(t))$ and $g(t) = O(f(t))$. A similar notion is used for sequences.

3 Examples

We would now like to see what the theorem gives by studying a few typical examples. As the remark following Lemma 2 indicates the techniques used in the proof are techniques that live in L_1. In most instances inequality (3) can be improved by using the estimate (11) in terms of $\alpha_{n,p}$ (Example 1) or an interpolation argument (Example 2). However, the Examples 4 and 5 show that in general the estimates (3) and (4) cannot be improved.

1. On $L_p[0,1]$, $1 \le p \le \infty$, we define for each $n = 1,2,\ldots$ the transformation $f \to E_n f$ as the conditional expectation of f with respect to the field generated by the set $\{[\frac{k}{n+1}, \frac{k+1}{n+1}) : k = 0,1,\ldots,n\}$, i.e.,

$$E_n f(x) = (n+1) \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} f(u)du, \quad x \in [\frac{k}{n+1}, \frac{k+1}{n+1}).$$

By definition, $(E_n : n = 1,2,\ldots)$ is a sequence of positive linear contractions, $E_n^2 = E_n$ and $E_n\phi_o \equiv \phi_o$. We have $\lambda_{n,p}^2 = \frac{1}{2}(p+1)^{-1/p}(n+1)^{-1}$. One verifies easily, that $\alpha_{n,p} = \lambda_{n,p}$ and, consequently, by (11) for any $f \in L^p[0,1]$

$$\|E_n(f) - f\|_p \le C\{\frac{1}{n}\omega_{1,p}(f;1) + \omega_{2,p}(f;n^{-1/2})\},$$

$$n = 1,2,\ldots$$

2. Bernstein-Kantorovič polynomials. For $f \in L_p[0,1]$ and $n = 1,2,\ldots$ let

$$P_n f(x) = B_n E_n f(x) = \frac{d}{dx} B_{n+1}F(x), \; F(x) = \int_o^x f(u)du,$$

where B_n is the well-known Bernstein operator. The P_n are positive linear contractions on $L_p[0,1]$ with $P_n\phi_o \equiv \phi_o$. Here $\lambda_{n,p}^2 \asymp (n+1)^{-1}$. For $p = 1$ this leads to

(12) $\|P_n f - f\|_p \le C_p\{\frac{1}{n}\omega_{1,p}(f;1) + \omega_{2,p}(f;n^{-1/2})\}$.

$f \in L_p[0,1], \quad n = 1,2,\ldots$

By Freud's estimate (1), this inequality also holds true for $p = \infty$. An interpolation argument finally establishes (12) for all p, $1 \le p \le \infty$.

The inequality improves estimates given by R. Bojanic and O. Shisha [2]. For functions f of bounded varation, (12) gives an order of approximation $n^{-1/2}$. In this case, a stronger estimate was given by W. Hoeffding [6] who included pointwise behavior at the endpoints.

3. For each t, $0 < t \le 1/2$ we define on $L_p[0,1]$, $1 \le p \le \infty$

$$L_t f(x) = \begin{cases} \frac{1}{2}(f(x+t) + f(x)) , & 0 \le x < t, \\ \frac{1}{2}(f(x+t) + f(x-t)), & t \le x \le 1-t, \\ \frac{1}{2}(f(x) + f(x-t)) , & 1-t < x \le 1. \end{cases}$$

Clearly, L_t defines a positive linear contraction on $L_p[0,1]$, and $L_t\phi_o \equiv \phi_o$. Also $\lambda^2_{t,p} \asymp t^{1+1/p}$. For any $f \in L_1[0,1]$ and any $0 < \delta \le 1/2$

$$\omega_{2,1}(f;\delta) \le \sup_{0<t<\delta} \|L_t f - f\|_1 \le C_1\{\delta^2\omega_{1,1}(f;1)+\omega_{2,1}(f;\delta)\};$$

with the right hand estimate derived from Theorem 1. It follows by the remark after Lemma 1 that the Lipschitz classes

$$\mathrm{Lip}^*(\alpha;L_1[0,1]) = \{f \in L_1[0,1] : \omega_{2,1}(f;\delta) \le C_f\delta^\alpha, \quad 0 < \delta \le 1/2,\ 0 < \alpha \le 2\},$$

are vector sublattices in $L_1[0,1]$. This statement is trivial for the class $\mathrm{Lip}(\alpha;L_p[0,1])$, $0 < \alpha \le 1$, for all $1 \le p \le \infty$, but for $\mathrm{Lip}^*(\alpha;L_1[0,1])$, $0 < \alpha \le 2$, it seems not so obvious. For $L_p[0,1]$, $1 < p \le \infty$, the latter is obviously wrong.

4. This example will show that (4) cannot be improved. For $f \in L_p[-1,1]$, $1 \leq p \leq \infty$, set

$$L_n f(x) = \begin{cases} f(x), & |x| \leq 1 - \frac{1}{n}, \\ f(-x), & 1-\frac{1}{n} < |x| \leq 1. \end{cases} \quad n = 1,2,\ldots$$

L_n forms a positive linear contraction on $L_p[-1,1]$ and $L_n\phi_o \equiv \phi_o$. Furthermore, $\lambda^2_{n,p} \asymp n^{-1/p} \asymp \lambda^{2/p}_{n,1}$. The function $f_o(x) = x^2_+$ belongs to $W_{2,p}[-1,1]$, but $\|L_n f_o - f_o\|_p \asymp n^{-1/p}$.

5. Finally, consider on $L_p[-1,1]$ the sequence of operators

$$L_n f(x) = \frac{1}{1+2^{-n}} \begin{cases} f(x), & 2^{-n} \leq |x| \leq 1, \\ \frac{1}{2}\int_{-1}^{1} f(x)dx, & 0 \leq |x| < 2^{-n} \end{cases} \quad n = 1,2,\ldots$$

One verifies easily that L_n defines a positive linear contraction on $L_p[-1,1]$, $1 \leq p \leq \infty$. Also $(1+2^{-n})L_n\phi_o(x) \equiv \phi_o(x)$ and

$$(1+2^{-n})L_n\phi_1(x) = \begin{cases} \phi_1(x), & 2^{-n} < |x| < 1, \\ 0, & |x| < 2^{-n}, \end{cases}$$

which implies that $\lambda^2_{n,p} \asymp 2^{-n}$.

For ϕ_2, we have $(1+2^{-n})L_n\phi_2(x) = \phi_2(x)$, $2^{-n} \leq |x| \leq 1$, and $= 1/3$ for $0 \leq |x| < 2^{-n}$. Hence

$$\|L_n\phi_2 - \phi_2\|_p \asymp \begin{cases} 2^{-n/p}, & 1 \leq p < \infty, \\ 1, & p = \infty, \end{cases}$$

proving that for sequences of positive linear contractions on $L_p[-1,1]$, $1 \leq p < \infty$, in general, the order of convergence given by (3) cannot be improved.

References

1 Berens, H. and G.G. Lorentz, Theorems of Korovkin type for positive linear operators on Banach lattices, in Approximation Theory, ed. by G.G. Lorentz, Academic Press, New York, 1973.

2 Bojanic, R. and O. Shisha, Degree of L_1 approximation to integrable functions by modified Bernstein polynomials, J. Approximation Theory 13 (1975), 66-72.

3 DeVore, R.A., The Approximation of Continuous Functions by Positive Linear Operators, Lecture Notes in Mathematics No. 293, Springer-Verlag, Berlin 1972.

4 ―――――――, Degree of approximation, these proceedings.

5 Freud, G., On approximation by positive linear methods, I and II, Studia Sci. Math. Hung. 2 (1967), 63-66, and 3 (1968), 365-370.

6 Hoeffding, W., The L_1 norm of the approximation error for Bernstein-type polynomials, J. Approximation Theory 4 (1971), 347-356.

7 Johnen, H., Inequalities connected with moduli of smoothness, Mat. Vesnik 9 (24) S. 3 (1972), 289-303.

H. Berens
Mathematisches Institut
Universität Erlangen-Nürnberg
8520 Erlangen, W. Germany

R. DeVore+
Department of Mathematics
Oakland University
Rochester, MI 48063
and
Mathematisches Institut
Universität Erlangen-Nürnberg
8520 Erlangen, W. Germany

+) Supported in part by the National Science Foundation and the Alexander von Humboldt-Stiftung.

REFLEXIVITY AND THE EXISTENCE OF BEST APPROXIMATIONS

Jörg Blatter

A (nonempty) subset A of a (real or complex) normed linear space X is called proximinal (or an existence subset) if every element of X has at least one best approximation in A. Problems in best approximation are especially simple in reflexive Banach spaces because in such spaces every closed convex subset is proximinal (this is a simple consequence of the weak compactness of the unit ball). It is thus natural to ask if there is a broader class of normed linear spaces in which every closed convex subset is proximinal. For Banach spaces this question is answered negatively by the following theorem of R. C. James [1]: A Banach space X is reflexive if and only if every continuous linear functional on X attains its norm on the unit ball. Namely, (as observed by James [3, p.253]) a continuous linear functional on a normed linear space attains its norm on the unit ball if and only if its kernel is proximinal. In 1966, F. Deutsch, in a letter to James, asked whether a normed linear space in which every closed convex subset is proximinal must be complete. After a couple of weeks James provided a partial answer by exhibiting an incomplete normed linear space in which every closed hyperplane is proximinal (much later this result was published in [2]). Despite this indication of a negative answer, Deutsch's question does indeed have a positive answer:

THEOREM. A normed linear space in which every closed convex subset is proximinal is complete.

Proof. Let X be an incomplete normed linear space. We shall exhibit a closed convex subset K of X in which 0

has no best approximation. Let $\hat{x}$ be a point of norm 1 in the completion of X which does not belong to X, and let $x' \varepsilon X'$ be of norm 1 and such that $x'\hat{x} = 1$. One easily constructs a sequence $\{x_n\}_{n\varepsilon\mathbb{N}}$ in X which converges to $\hat{x}$ and is such that $x'x_n = 1 + 1/n$ for all n. Set $K = \overline{co}(\{x_n : n \varepsilon \mathbb{N}\})$ (= closed convex hull of $\{x_n : n \varepsilon \mathbb{N}\}$); x' is real on K and for each $x \varepsilon K$, $x'x \geq 1$. Since $\|x'\| = 1$, $\|x\| \geq 1$ for all $x \varepsilon K$. $\lim_n \|x_n\| = \|\hat{x}\| = 1$ and so $\inf\{\|x\| : x \varepsilon K\} = 1$, i.e., the distance of 0 to K is 1. We shall thus be done once it is shown that $x'x > 1$ for $x \varepsilon K$. Accordingly, let us suppose that there is an $x \varepsilon K$ such that $x'x = 1$. We have obviously $\bigcap_{n\varepsilon\mathbb{N}} \overline{co}(\{x_k : k > n\}) = \emptyset$ and so there is an $N \varepsilon \mathbb{N}$ such that $x \notin \overline{co}(\{x_k : k > N\})$. Let $\{y_n\}_{n\varepsilon\mathbb{N}}$ be a sequence in $co(\{x_n : n \varepsilon \mathbb{N}\})$ (= convex hull of $\{x_n : n \in \mathbb{N}\}$) which converges to x. Set $A = co(\{x_k : k \leq N\})$ and $B = co(\{x_k : k > N\})$. We can find $a_n \varepsilon A$, $b_n \varepsilon B$ and nonnegative reals α_n and β_n such that $y_n = \alpha_n a_n + \beta_n b_n$ and $\alpha_n + \beta_n = 1$. For $a \varepsilon A$, $x'a \geq 1 + 1/N$ and for $b \varepsilon B$, $x'b \geq 1$. Thus

$$x'y_n = \alpha_n x'a_n + \beta_n x'b_n \geq \alpha_n(1 + \frac{1}{N}) + \beta_n = 1 + \frac{1}{N}\alpha_n .$$

Since $\{x'y_n\}_{n\varepsilon\mathbb{N}}$ converges to $x'x = 1$, and since

$$1 + \frac{1}{N}\alpha_n \geq 1$$

for all n, $\{\alpha_n\}_{n\varepsilon\mathbb{N}}$ must converge to 0. Since A is bounded, $\{\alpha_n a_n\}_{n\varepsilon\mathbb{N}}$ converges to 0, i.e., $\{\beta_n b_n\}_{n\varepsilon\mathbb{N}}$ converges to x. Since $\{\beta_n\}_{n\varepsilon\mathbb{N}}$ converges to 1, $\{b_n\}_{n\varepsilon\mathbb{N}}$ also converges to x, i.e., $x \varepsilon \overline{co}(\{x_k : k > N\})$--a contradiction.

This theorem, of course, combines with James' theorem to yield the following characterization of reflexivity: <u>a normed linear space is reflexive if and only if each of its closed</u>

convex subsets is proximinal. Because James' theorem is comparatively deep, it would be nice to have a simple direct proof of this characterization.

References

1 James, R.C., Characterizations of reflexivity. Studia Math. 23 (1964), 205-216.

2 __________, A counterexample for a sup theorem in normed spaces. Israel J. Math. 9 (1971), 511-512.

3 Phelps, R. R., Uniqueness of Hahn-Banach extensions and unique best approximation, Trans. Amer. Math. Soc. 95 (1960), 238-255.

Jörg Blatter
Instituto de Matemática
Pura e Aplicada
Rio de Janeiro, Brazil

APPROXIMATE FUNCTIONAL COMPLEXITY

R. Creighton Buck

Necessary conditions are obtained for a function F either to be representable in the form $F(x,y,z) = f(\phi(x,y),z)$ with ϕ continuous, or to be uniformably approximable by such functions.

The works of Vitushkin, Arnold, Kolmogorov, and others have answered many of the basic questions in the theory of functional representation and complexity ([1], [6], [8], [9]). However, these methods often depend on category arguments that do not seem to be suitable for deciding the status of an individual function (see [4], [5], but also [3]). It is of interest, therefore, to find criteria which help to decide if a specific function either belongs to, or can be approximated by, a given class of simple functions of a prescribed superposition format (see [2]). In this paper we illustrate this, using as a test class the functions of the form $F(x,y,z) = f(\phi(x,y),z)$.

We write $\mathcal{F}^n(\mathcal{O})$ for the class of such F which have this representation on the region $\mathcal{O}$, with component functions f and ϕ of class C^n. If $n \geq 2$, a characterization is immediate.

THEOREM 1. *If* $F \in \mathcal{F}^2(\mathcal{O})$, *then* F *satisfies in* $\mathcal{O}$ *the nonlinear equation*

$$(1) \quad F_x F_{yz} - F_y F_{xz} = 0$$

Preceding page blank

Conversely, if F satisfies this equation in $\mathcal{O}$, then F has the desired local representation in the set $\mathcal{O} - \Gamma$, where Γ is the set where $F_x = F_y = 0$.

One would hope for a similar characterization with only continuity requirements on f and ϕ, perhaps a weak form of (1). As a first step toward this, observe that Theorem 1 can be given a different formulation. For any c, consider the planar mapping T defined by

$$(2) \quad T : \begin{cases} u = F(x,y,c) \\ v = F_z(x,y,c) \end{cases}$$

Then, equation (1) can be restated as $\partial(u,v)/\partial(x,y) = 0$. Thus, $F \in \mathcal{F}^2$ if the mapping T is locally singular, for each c. With this, one is led to a C^1 characterization, based on the observation that F_z is approximately $F(x,y,z+h) - F(x,y,z)$.

THEOREM 2. Define a planar mapping by

$$(3) \quad T : \begin{cases} u = F(x,y,c_1) \\ v = F(x,y,c_2) \end{cases}$$

Then $\partial(u,v)/\partial(x,y) = 0$ for all (x,y,c_i) in $\mathcal{O}$ if $F \in \mathcal{F}^1(\mathcal{O})$. Conversely, if this holds, F is locally of the desired form in $\mathcal{O} - \Gamma$.

This is not yet the theorem sought, since it requires C^1 components, but it is suggestive. It is helpful, before going further, to recast the original problem. Given $F\colon X \times Y \times Z \to R$, of the form $f(\phi(x,y),z)$, an equivalent viewpoint is to consider F as a mapping from $X \times Y$ into $C[Z]$, which may now be replaced by any Banach space $\mathcal{E}$. The representation problem is now converted into a factoring problem; given F, find ϕ and f so that the following diagram commutes :

(4)
$$\begin{tikzcd} R^2 \arrow[rr, "\phi"'] \arrow[dr, "F"'] & & R \arrow[dl, "f"] \\ & \mathcal{E} & \end{tikzcd}$$

If no restrictions are placed on ϕ, this may be solved for arbitrary F by taking ϕ as any 1:1 map from R^2 into R. If we ask that ϕ be continuous, F is severely restricted.

THEOREM 3. *If* F *has the form shown in* (4) *and* ϕ *is continuous, then* F *is locally singular in the sense that on any open set, either* F *is constant or* F *has a noncountable number of distinct noncountable level sets (sets of constancy)*.

This elementary result is sufficient to show that specific functions do not belong to the class $\mathcal{F}^0$ on any open set (e.g., $xy + yz + xz$).

This result can now be extended to the uniform closure of $\mathcal{F}^0$ by means of a special property of continuous functions $\phi\colon R^2 \to R$.

LEMMA. *Let* S *be the unit square in* R^2. *There is a constant* C *such that for every continuous real function* ϕ *on* S, *and* $0 < \delta < 1$, *there is a level set* E *for* ϕ *such that* $N(\delta,E) > C/\delta$.

Here, $N(\delta,E)$ is a standard "size" function for subsets of a metric space, the minimum number of sets required for a δ cover for E. Using this Lemma, one may obtain a criterion for approximate representability.

THEOREM 4. *If* $F_o\colon S \to \mathcal{E}$ *is the uniform limit on* S *of functions* F *of the form* $f(\phi(x,y))$ *where* $\phi\colon R^2 \to R$ *is*

continuous, then F_o must have level sets of arbitrarily large finite cardinal.

This shows immediately that functions such as $xy + yz + xz$ or $x^2y + y^2z + z^2x$ do not lie in the uniform closure of F^0 on any open set. Indeed, a refinement of this approach allows one to obtain an estimate for the distance from F_o to F^0 when F_o is sufficiently smooth. These results suggest the following conjectures: (i) if $F_o \varepsilon C^\infty$, and F_o lies in the uniform closure of F^0, then F_o satisfies the differential equation (1); (ii) if $F_o \varepsilon C^\infty$, then the uniform norm distance between F_o and $F^0(S)$ can be estimated by the minimum of $|H(F_o)|$ where H is the differential operator in (1).

References

1 Arnold, V. I. Some questions on approximation and representation. Proc. Int. Cong. Math. Edinburgh, 1958. Amer. Math. Soc. Transl. 53 (1966), 192-210.

2 Buck, R. C., Approximate functional complexity, Bull Amer. Math. Soc. 81 (1975), 1112-1114.

3 Doss, R., On the representation of continuous functions of two variables, Colloq. Math. 10 (1963), 249-259.

4 Fridman, B. L., Nowhere denseness of the space of linear superpositions, Izv. Akad. Nauk SSSR, Ser. Mat. 36 (1972), 814-846.

5 Kaufman, R., Linear superpositions of smooth functions, Proc. Amer. Math. Soc. 46 (1974), 360-362.

6 Kolmogorov, A. N., On the representation of continuous functions of several variables, Amer. Math. Soc. Transl. 28 (1963).

7 Lorentz, G., Metric entropy, widths, and superposition of functions, Amer. Math. Monthly 69 (1962), 469-485.

8 Sprecher, D., A survey of solved and unsolved problems on superposition of functions. J. Approximation Theory 6 (1972), 123-134.

9 Vitushkin, A. G., Theory of the transmission and processing of information, Pergamon Press N.Y. 1961.

R. C. Buck
Dept. Mathematics
Univ. of Wisconsin
Madison, WI 53706

Supported in part by NSF grant MPS75-06687 and ARO contract DA31-124AROD462.

BEST UNIFORM HARMONIC APPROXIMATION

H. G. Burchard

For a given continuous real function on the closed unit disc, we characterize its best uniform harmonic approximation, if this exists.

Let Δ be the closed unit disc in the complex plane $\mathbb{C}$. $C(\Delta)$ is the Banach space of real continuous functions on Δ, with the uniform norm, H the closure in $C(\Delta)$ of the real harmonic functions on $\mathbb{C}$ restricted to Δ, i.e., H consists of the Poisson integrals of continuous data on $\mathrm{bdy}(\Delta) = T$.

If $f \in C(\Delta)$, then f need not have a best approximation in H. However, enlarge H to the set H^{∞} of Poisson integrals of L^{∞} - data on T. Then there exists $h \in H^{\infty}$ and a minimizing sequence (h_n) in H which converges uniformly on compact subsets of $\mathrm{int}(\Delta)$ to h, a best approximation from H^{∞} to f (in the space of bounded functions on Δ).

DEFINITION 1. *For a compact set* $K \subset \mathbb{C}$ *let* $\Omega(K)$ *be the unbounded component of* $\mathbb{C} - K$ *and* $K^{\wedge} = \mathbb{C} - \Omega(K)$.

$K^{\wedge}$ is the polynomially convex hull of K.

THEOREM 1. *If* $f \in C(\Delta)$, $g \in H$ *let*

$$K_{\pm} = \{x \in \Delta : f(x) - g(x) = \pm\|f - g\|_{C(\Delta)}\}.$$

Then g *is a best uniform approximation to* f *from* H *if and only if the following condition holds*:

(1) *It is not true that both* $K_{+} \subset \Omega(K_{-})$ *and* $K_{-} \subset \Omega(K_{+})$.

One can also express (1) by saying K_{+} meets $K_{-}^{\wedge}$ or K_{-} meets $K_{+}^{\wedge}$.

Preceding page blank

A remark of R. C. Buck in [5] suggested to the author that there should be such a theorem.

The proof of Theorem 1 is based on one half of the following known result, *cf*. [8].

THEOREM 2. *Given is a compact Hausdorff space* X *and a linear subspace* A *of (real)* $C(X)$. *Denote by* $A^{\perp}$ *the annihilator of* A *in* $C(X)^*$ *(*$A^{\perp}$ *= set of Radon measures orthogonal to* A*). If* $f \in C(X)$ *and* $g \in A$ *let*

$$K_{\pm} = \{x \in X: \; f(x) - g(x) = \pm||f - g||_{C(X)}\}.$$

Then g *is a best (uniform) approximation to* f *from* A *if and only if one of the following two equivalent conditions holds*:

(2) $\exists\ h \in A,\ h(K_{+}) > 0 > h(K_{-})$ *(elementwise)*;

(3) $\exists\ \mu \in A^{\perp},\ \mu \neq 0,\ \mathrm{supp}(\mu^{\pm}) \subset K_{\pm}$.

We also require the following intuitive result.

LEMMA 1. *Suppose* K_1, K_2 *are compact subsets of the plane,* $K_1 \subset \Omega(K_2)$, $K_2 \subset \Omega(K_1)$. *Let* $K_0 = K_1 \cup K_2$. *Then* $K_1^{\wedge} \cap K_2^{\wedge} = \emptyset$ *and* $K_0^{\wedge} = K_1^{\wedge} \cup K_2^{\wedge}$.

Proof. One first shows $K_1^{\wedge} \cap K_2^{\wedge} = \emptyset$ by distinguishing several cases. This shown, let $\Omega_1 = \Omega(K_1)$, $\Omega_2 = \Omega(K_2)$. Then $\Omega_1 \cup \Omega_2 = \mathbb{R}^2$ and each Ω_j is a connected open set. It follows that $A = \Omega_1 \cap \Omega_2$ is connected: By the Mayer-Vietoris Theorem there is an exact sequence

$$0 = H_1(\mathbb{R}^2) \to H_0(A) \to H_0(\Omega_1) \oplus H_0(\Omega_2) \to H_0(\mathbb{R}^2) = 0$$

using singular homology in the augmented case [2]. But Ω_1 and Ω_2 are path connected, hence $H_0(\Omega_1) = H_0(\Omega_2) = 0$, so $H_0(A) = 0$, i.e. $\Omega_1 \cap \Omega_2 = \mathbb{C} - (K_1^{\wedge} \cup K_2^{\wedge})$ is path connected. From this one deduces $K_0^{\wedge} = K_1^{\wedge} \cup K_2^{\wedge}$.

The following result now implies Theorem 1.

PROPOSITION 1. _Let_ K_1, K_2 _be compact subsets of_ Δ. _Then_ (1) _and_ (2), _with_ $A = H$, _are equivalent_.

Proof. Assume (1) fails. The conclusions of Lemma 1 hold and we let $u = 1$ on $K_1^\wedge$, $u = -1$ on $K_2^\wedge$. Then u is continuous on $K_0^\wedge$, which is compact with connected complement. By the Walsh-Lebesgue Lemma, we can find $h \varepsilon H$ such that $|h - u| < \frac{1}{2}$ on $\mathrm{bdy}(K_0^\wedge)$. By the maximum principle, $h(K_1) > 0 > h(K_2)$, so (2) fails (or use Runge's theorem in place of the Walsh-Lebesgue Lemma, [S. Minsker's remark]). Now suppose (2) fails, $A = H$, and $h \varepsilon H$ such that $h(K_1) > 0 > h(K_2)$. If $x \varepsilon K_2^\wedge$ then either $x \varepsilon K_2$, $h(x) < 0$, so $x \not\varepsilon K_1$, or $x \varepsilon \Omega$, a bounded component of $\mathbf{C} - K_2$. But $\mathrm{bdy}(\Omega) \subset K_2$, so $h(x) < 0$ by the maximum principle. Hence $x \not\varepsilon K_1$. This shows $K_1 \subset \Omega(K_2)$ and thus also $K_2 \subset \Omega(K_1)$, i.e., (1) fails.

The same method of proof gives the analogous result for harmonic functions on the unitball of R^n; for generalizations of the Walsh-Lebesgue Lemma (or Runge's theorem) _cf_. [4].

PROBLEM. Replace harmonic functions by biharmonic functions to obtain a subspace BH of $C(\Delta)$. SH is the convex cone of continuous subharmonic functions on Δ. Prove results similar to Theorem 1 for BH and SH. The latter case is analogous to problems on the line solved in [6].

The known generalizations of the Walsh-Lebesgue Lemma to solutions of elliptic equations [4] do not appear to be sufficient to solve the problem for BH. E.g., one can show that $BH|_K$ is dense in $C(K)$ if K is the union of two concentric circles. One of these may be deformed into a rectifiable Jordan curve.

It is of interest to characterize further the sets K^+ that

can occur in (2) or (3). Let

$$\Sigma_2 = \{K \subset X: \ K \text{ is closed}, \exists \mu \in A^{\perp}, \mu \neq 0, \ \mathrm{supp}\, \mu^{+} \subset K\},$$

$$\Sigma_0 = \{K \in \Sigma_2: \ \exists \mu \in A^{\perp}, \mu \neq 0, \ \mathrm{supp}\, \mu^{+} \subset K, \ K \cap \mathrm{supp}\, \mu^{-} = \emptyset\}.$$

$$\Sigma_1 = \{K \in \Sigma_2: \exists K_1 \subset K, \ K_1 \in \Sigma_0\}.$$

Then $\Sigma_0 \subset \Sigma_1 \subset \Sigma_2$.

PROPOSITION 2. *For* $A = H$, *and* $K \subset \Delta$, K *closed, one has* $K \in \Sigma_2 = \Sigma_1$ *iff* K *is not a proper subset of* T; $K \in \Sigma_0$ *iff* $K \cap T = \emptyset$ *or* $K \neq K^{\wedge}$.

One can prove a quite general theorem characterizing Σ_2 using Choquet Theory. A is as above, but in addition $1 \in A$ and A separates the points of X.

THEOREM 3. *For a closed* $K \subset X$ *each of the following is equivalent with* $K \notin \Sigma_2$:

(4) *If* $\mu \in A^{\perp}$, $\mathrm{supp}\, \mu^{+} \subset K$, *then* $\mu = 0$.

(5) a) *If* $\mu \in A^{\perp}$, $\mathrm{supp}\, \mu^{+} \subset K$, *then* $\mathrm{supp}\, \mu^{-} \subset K$.

b) $A|_K$ *is dense in* $C(K)$.

(6) *If* $\mu \in M_1^{+}(X)$, $\mathrm{supp}\, \mu \subset K$, *then* μ *is the unique representing measure of* $\mu|_A$.

(7) *If* $\mu \in M_1^{+}(X)$, $\mathrm{supp}\, \mu \subset K$, *then* $\forall f$ *in* $C(X)$

$\int f\mu = \sup\{\int g\mu: \ g \in A, \ g \leq f\}$.

(8) a) *The barycenters (in the state space* S *of* A*) of probability measures carried by* K *form an extremal Bauer simplex (extreme boundary is closed* [1]) $\mathrm{bary}(K) \subset S$ *(we let canonically* $K \subset X \subset S$, $A \subset A(S)$).

b) *If* $\varepsilon_x|_A \in \mathrm{bary}(K)$, *then* $x \in K$ (ε_x = *unit point mass*).

c) $\{\varepsilon_x : x \in K\} = \mathrm{extr}(\mathrm{bary}(K)) \subset \mathrm{extr}(S)$.

(9) (5)b) holds and with $T = \mathrm{bary}(K)$

$\forall\ \alpha, \beta,\ 0 < \alpha < \beta,\ \forall$ open $V \supset K$, $\exists\ f_1, \ldots, f_n \in A$,

$0 \geqq f_i$, $\forall x \in T \ \max\limits_{1 \leqq i \leqq n} f_i(x) \geqq -\alpha > -\beta \geqq f_j(X - V)$.

If $K = \{x\}$, then $K \notin \Sigma_2$ if $x \in \partial_A(X)$, the Choquet boundary of A. The proof of Theorem 3 mostly parallels similar proofs for $K = \{x\}$, cf. [3]. Much of Theorem 3 can be extended to arbitrary Borel subsets K of X.

For $A = BH$ one might hope that $\Sigma_2 = \Sigma_1$, too. Note that $\mu \in BH^{\perp}$ iff $\mu \in H^{\perp}$ and $r^2\mu \in H^{\perp}$ [7]. Also, $\partial_{\Delta}BH = \Delta$.

References

1 Alfsen, E. M., Compact Convex Sets And Boundary Integrals, Springer Verlag, New York 1971.

2 Artin E., and H. Braun, Introduction to Algebraic Topology, Ch. E. Merrill, Columbus 1969.

3 Browder, A., Introduction to Function Algebras, Benjamin, New York 1969.

4 Browder, F. E., Approximation by solutions of partial differential equations, Amer. J. Math., 84(1962), 134-160.

5 Buck, R. C., Applications of duality in approximation theory, in: Approximation of Functions, H. L. Garabedian, ed., Elsevier, Amsterdam 1965, 27-42.

6 Burchard, H. G., Extremal positive splines and applications, in: Approximation Theory, G. G. Lorentz, ed., Academic Press, New York 1973, 291-294.

7 Nicolesco, M., Les Fonctions Polyharmoniques, Actual. Sc. Ind. 331, Hermann & Cie, Paris 1936.

8 Singer, I., Best Approximation in Normed Linear Spaces, Springer Verlag, New York 1970.

H. G. Burchard
Department of Mathematics
Oklahoma State University
Stillwater, Oklahoma 74074

Research supported in part by Grant MPS-74-06995-A01 of the National Science Foundation.

FRACTIONAL CHEBYSHEV OPERATIONAL CALCULUS AND BEST ALGEBRAIC APPROXIMATION

P.L. Butzer and R.L. Stens

A basic difficulty in the fundamental theorem of best approximation of functions defined on a finite interval $[-1,1]$, say, by algebraic polynomials, is due to the fact that approximation is more efficient near the end points $-1,+1$ than within the interval. What is moreover lacking seems to be a systematic approach, made possible in the trigonometric case by the finite Fourier transform. To overcome these and other difficulties, the purpose of this paper is to present a new approach to this subject based upon a Chebyshev operational calculus.

1 Introduction

The Chebyshev transform of $f \in X$, where $X = C[-1,1] \equiv C$ or $L^p_w(-1,1) \equiv L^p_w$, $1 \leq p < \infty$, with norms

$$\|f\|_C = \sup_{x \in [-1,1]} |f(x)|, \quad \|f\|_p = \{\frac{1}{\pi}\int_{-1}^{1} |f(u)|^p \, w(u)du\}^{1/p}$$

respectively, where $w(x) = (1-x^2)^{-1/2}$, is defined by

$$f^\wedge(k) = \frac{1}{\pi}\int_{-1}^{1} f(u)T_k(u)w(u)du \qquad (k \in P),$$

$T_k(x) = \cos(k \arccos x)$, $x \in [-1,1]$, being the Chebyshev polynomial of degree k. The associated translation operator for $f \in X$ is defined for $x, h \in [-1,1]$ by

$$(\tau_h f)(x) = \frac{1}{2}\{f(xh + \sqrt{(1-x^2)(1-h^2)}) + f(xh - \sqrt{(1-x^2)(1-h^2)})\}$$

and the convolution product of $f \in X$, $g \in L^1_w(-1,1)$ by

$$(f*g)(x) = \frac{1}{\pi}\int_{-1}^{1} (\tau_x f)(u)g(u)w(u)du \qquad (x \in [-1,1]).$$

2 Fractional Chebyshev Derivatives and Integrals

Defining the (right) difference of $f \in X$ of arbitrary order

$\alpha>0$ with respect to the increment $h\in[-1,1]$ by

$$(\overline{\Delta}_h^{\alpha} f)(x) := (-1)^{[\alpha]} \sum_{j=0}^{\infty} (-1)^j \binom{\alpha}{j} (\tau_h^j f)(x) \qquad (x\in[-1,1]),$$

where $[\alpha]$ is the largest integer $\leqslant\alpha$, and the strong Chebyshev derivative of $f\in X$ of order α as the function $g\in X$ for which

$$\lim_{h\to 1-} \left\| \frac{\overline{\Delta}_h^{\alpha} f}{(1-h)^{\alpha}} - g \right\|_X = 0;$$

we write $D^{\alpha}f=g$ if such g exists. The Chebyshev integral $I^{\alpha}f$ of f of order $\alpha>0$ is defined by

$$(I^{\alpha}f)(x) = (f * \psi_{\alpha})(x) \qquad (x\in[-1,1]),$$

where ψ_{α} is the L_w^1-function defined uniquely by

$$\psi_{\alpha}^{\wedge}(k) = \begin{cases} (-1)^{[\alpha]} k^{-2\alpha}, & k\in\mathbb{N} \\ 0 & , \ k=0. \end{cases}$$

THEOREM 1. *The following assertions are equivalent for* $f\in X, \alpha>0$:

(i) $D^{\alpha}f$ *exists as an element of* X,

(ii) *there exists a function* $g_0\in X$ *such that*

$$(-1)^{[\alpha]} k^{2\alpha} f^{\wedge}(k) = g_0^{\wedge}(k) \qquad (k\in\mathbb{N}),$$

(iii) *there exists* $g_1\in X$ *such that* $f(x) = (I^{\alpha}g_1)(x) + \text{const.}$ (a.e.).

The functions g_i *are uniquely determined (a.e.) apart from an additive constant, and one has* $(D^{\alpha}f)(x) = g_i(x) - g_i^{\wedge}(0)$ (a.e.).

Setting $W_X^{\alpha} := \{f\in X;\ D^{\alpha}f \text{ exists in } X\}$, one has as immediate consequences of Thm. 1 that the operator $D^{\alpha}:W_X^{\alpha}\to X$ is closed, and the "fundamental theorem for fractional Chebyshev derivatives and integrals":

$$D^{\alpha}(I^{\alpha}f) = f - f^{\wedge}(0) = I^{\alpha}(D^{\alpha}f) \qquad \text{(a.e.)},$$

the latter equality holding provided $f\in W_X^{\alpha}$. Similarly D^{α} satisfies the additivity law. This shows that $D^1 f$ can be defined equivalently by the limit in norm of X of $(\tau_h f-f)/(1-h)$ for $h\to 1-$, and for $r=2,3,\dots$ by $D^r f = D^1(D^{r-1}f)$.

3 Pointwise Interpretations

In order to express the strong Chebyshev derivative of order $\alpha>0$ in terms of the usual pointwise derivatives, we need to define a conjugate function to $f\in X$, given by

$$f^{\sim}(x) = \lim_{r\to 1-} 2\sum_{k=1}^{\infty} r^k f^{\wedge}(k)\sin(k \arccos x) \qquad (x\in[-1,1]),$$

whenever the limit exists. Restricting the matter to the simpler case $X=C[-1,1]$ (for $L^p_w(-1,1)$ see [2,4]) one has

THEOREM 2. *If* $f\in W^{\alpha}_{C}$, *then one has for each* $r\in\mathbb{N}$, $r\geq\alpha$ *and each* $s\in\mathbb{P}$, $2s+1\geq\alpha$ *for* $x\in(-1,1)$

$$(D^{\alpha}f)(x) = (-1)^{r+[\alpha]+[r-\alpha]}\left[(1-x^2)\frac{d^2}{dx^2} - x\frac{d}{dx}\right]^r (I^{r-\alpha}f)(x),$$

$$(D^{\alpha}f)(x) = (-1)^{t(s,\alpha)}\sqrt{1-x^2}\,\frac{d}{dx}\left(\left[(1-x^2)\frac{d^2}{dx^2} - x\frac{d}{dx}\right]^s (I^{s-\alpha+1/2}f)\right)^{\sim}(x),$$

with $t(s,\alpha)=s+1+[\alpha]+[s-\alpha+1/2]$, $I^0f=f$. *The cases* $\alpha=1$, $\alpha=1/2$ *read*

$$(D^1f)(x) = (1-x^2)f''(x) - xf'(x) \qquad (x\in(-1,1)),$$

$$(D^{1/2}f)(x) = -\sqrt{1-x^2}\,(f^{\sim})'(x) \qquad (x\in(-1,1)).$$

The assertions remain valid for $x=\pm 1$ *if the limits* $x\to 1-$ *and* $x\to(-1)+$ *are taken on the right hand sides.*

4 Lipschitz Classes and Best Approximation

Defining the Chebyshev modulus of continuity of $f\in X$ of order $\alpha>0$ by

$$\omega^T_{\alpha}(X;f;\eta) = \sup_{\eta\leq h\leq 1} \|\bar{\Delta}^{\alpha}_h f\|_X \qquad (\eta\in[-1,1]),$$

the associated Lipschitz class of order $\gamma>0$ is given by

$$\mathrm{Lip}^T_{\alpha}(X;\gamma) = \{f\in X;\ \omega^T_{\alpha}(X;f;\eta) = O((1-\eta)^{\gamma}),\ \eta\to 1-\}.$$

We denote the best approximation of $f\in X$ by algebraic polynomials $p_n(x)$ of degree $\leq n$ by $E_n(X;f)$. It is well known that there exists a $p^*_n(x) = p^*_n(x;f)$ such that $E_n(X;f) = \|f-p^*_n\|_X$.

Now to the generalization of the fundamental theorem on best algebraic approximation to derivatives and moduli of continuity of fractional order.

THEOREM 3. *The following assertions are equivalent for* $f\in X(=C[-1,1]$ *or* $L^p_w)$ *and* $\alpha_2<\alpha+\beta<\alpha_1$, $\alpha,\alpha_2\geq 0$, $\alpha_1>0$, $0<\beta\leq 1$:

(i) $E_n(X;f) = O(n^{-2(\alpha+\beta)})$ $(n\to\infty)$,

(ii) $\omega^T_{\alpha_1}(X;f;\eta) = O((1-\eta)^{\alpha+\beta})$ $(\eta\to 1-)$, *i.e.* $f\in \mathrm{Lip}^T_{\alpha_1}(X;\alpha+\beta)$,

(iii) $D^{\alpha}f \in \begin{cases} \mathrm{Lip}^T_1(X;\beta), & 0<\beta<1 \\ \mathrm{Lip}^T_2(X;\beta), & \beta=1, \end{cases}$

(iv) $\|D^{\alpha_1}p^*_n\|_X = O(n^{-2(\alpha+\beta-\alpha_1)})$ $(n\to\infty)$,

(v) $f\in W^{\alpha_2}_X$ *and* $\|D^{\alpha_2}f-D^{\alpha_2}p^*_n\|_X = O(n^{-2(\alpha+\beta-\alpha_2)})$ $(n\to\infty)$.

Let us just compare our implication (iii) $\Rightarrow$ (i) for $\alpha=r-1/2$, $r\in\mathbb{N}$, and $X=C[-1,1]$, namely that

$$(1)\quad D^{r-1/2}f\in \mathrm{Lip}^T_1(C;\beta/2) \Rightarrow E_n(C;f) = O(n^{-2r+1-\beta}),$$

with the classical result of Timan [8], Teljakowskiĭ [7], Gopengauz [6]: to $f^{(\alpha)}\in C[-1,1]$, $\alpha\in\mathbb{N}$, there exists a sequence of algebraic polynomials such that

$$|f(x)-p_n(x)| = O\Big(\big(\tfrac{1}{n}\sqrt{1-x^2}\big)^{\alpha}\,\Omega\big(f^{(\alpha)};\tfrac{\sqrt{1-x^2}}{n}\big)\Big),$$

$$\Omega(f;\delta) := \sup\{|f(x)-f(y)|;\ x,y\in[-1,1],\ |x-y|<\delta\}.$$

Neglecting the factors $\sqrt{1-x^2}$ on the right, this result states in particular that ($\mathrm{Lip}_1(C;\beta)$ = usual Lipschitz class)

$$(2)\quad f^{(2r-1)}\in \mathrm{Lip}_1(C;\beta),\ r\in\mathbb{N},\ 0<\beta\leq 1 \Rightarrow E_n(C;f) = O(n^{-2r+1-\beta}).$$

Comparing the hypotheses in (1) and (2), ours, equivalent to $\sqrt{1-x^2}\, d/dx\big([(1-x^2)d^2/dx^2-x\, d/dx]^{r-1}f\big)^{\sim}(x)\in \mathrm{Lip}^T_1(C;\beta/2)$ by Thm. 2, is truly weaker than $f^{(2r-1)}\in \mathrm{Lip}_1(C;\beta)$. Here the multiplicative factor $(\sqrt{1-x^2})^{2r-1}$ of the highest order derivative compensates

the factors dropped above.

For other results that have some connections to particular cases of ours see [5]. The proofs of our results as well as a number of others will follow in [1-4].

References

1 Butzer, P.L. and R.L. Stens, The operational properties of the Chebyshev transform. I: General properties, Funct. Approximatio Comment. Math. 5 (1977) (to appear).

2 Butzer, P.L. and R.L. Stens, The operational properties of the Chebyshev transform. II: Fractional derivatives, in: Proc. Int. Conf. on Theory of Approximation of Functions, July 1975, Kalouga, USSR (to appear).

3 Butzer, P.L. and R.L. Stens, Chebyshev transform methods in the theory of best algebraic approximation, Abh. Math. Sem. Univ. Hamburg 45 (1976) (to appear).

4 Butzer, P.L. and R.L. Stens, The fractional Chebyshev operational calculus and the fundamental theorems on the approximation of functions by algebraic polynomials (in preparation).

5 Fuksman, A.L., Structural characteristic of functions such that $E_n(f;-1,1) \leq Mn^{-(k+\alpha)}$ (Russian), Uspehi Mat. Nauk 20 (1965), 187-190.

6 Gopengauz, I.E., A question concerning the approximation of functions on a segment and a region with corners (Russian), Teor. Funkciĭ Funkcional. Anal. i Priložen. 4 (1967), 204-210.

7 Teljakowskiĭ, S.A., Two theorems on the approximation of functions by algebraic polynomials (Russian), Mat. Sb. (N.S.) 70 (1966), 252-265 = Amer. Math. Soc. Transl. 77 (1968), 163-178.

8 Timan, A.F., Theory of Approximation of Functions of a Real Variable, Pergamon Press, Oxford 1963.

P.L. Butzer, R.L. Stens
Lehrstuhl A für Mathematik
Technological University of Aachen
Templergraben 55
51 Aachen, BRD

ON THE COMPUTATION OF MINIMAL PROJECTIONS FROM $C[0,1]$ TO $P_n[0,1]$

B. L. Chalmers and F. T. Metcalf

This paper discusses the computational implications of several theoretical developments concerning minimal projections from $C[0,1]$ to $P_n[0,1]$ (the polynomials of degree at most n). Utilizing these ideas, several cases have been numerically treated for projections to the quadratics, and the results are given below.

1 Introduction

The "minimal projection problem," for projections from $C[0,1]$ to $P_n[0,1]$, has generated a large body of literature, very little of which will be cited here. This problem, though simple to state, has been settled only for the case $n = 1$, or for certain subclasses of projections (see [1]). Throughout what follows, the projections will be assumed to be "finite carrier projections;" that is, in the representation

$$P = \Sigma_{i=0}^{n} L_i \otimes v_i; \quad L_i \in C^*, \ v_i \in P_n, \ L_i v_j = \delta_{ij},$$

the functionals L_i are carried on a finite number of points. It is known that any projection is approximable by finite carrier projections (see [2]).

The determination of the projection constant (minimal norm) consists of the min-max-max problem

$$\min_{L_i \in C^*} \ \max_{\substack{f \in C[0,1] \\ \|f\|_\infty = 1}} \ \max_{x \in [0,1]} |(Pf)(x)|.$$

Computationally, the min/max operations may be viewed as

Preceding page blank

follows, where it is assumed that the projection is carried on m points.

(I) max over $x \in [0,1]$--Finding the extrema of an $n^{\underline{th}}$ degree polynomial.

(II) max over $f \in C[0,1]$, $\|f\|_\infty = 1$--Maximizing over the 2^m possible ± 1 values for f at the m carriers.

(III) min over $L_i \in C^*$--(a) Minimizing over the $m(n+1)$ coefficients in the expansion of the L_i by point evaluations $(i = 0,\ldots,n)$. (b) Minimizing over the choices for the m carriers.

Problem (III) presents the major difficulty in the form of $m(n + 1)$ linear parameters (the coefficients) and m non-linear parameters (the carriers). The goal of what follows is to reduce the number of parameters in (III), and to a lesser extent, the number of choices in (II).

An immediate reduction in the number of parameters in (IIIa) may be obtained as follows. First, note that P is invariant under any invertible linear transformation of the functionals $L = (L_o,\ldots,L_n)$; i.e.,

$$P = L \otimes v = (LA) \otimes (vA^{t^{-1}})$$

for any nonsingular $(n + 1)x(n + 1)$ matrix A. Thus, the matrix of coefficients of the functionals L_i may be placed in echelon form, and $(n + 1)^2$ of the coefficient parameters eliminated. This gives, in place of (IIIa):

(IIIa') Minimizing over $(m - n - 1)(n + 1)$ coefficients in the expansion of the L_i by point evaluations.

Also, the number of carriers can be reduced by two, so that (IIIb) becomes

(IIIb') Minimizing over $m - 2$ choices for the carriers.

This follows from the fact that amongst the minimal projections on m carriers, there will be a projection having the

end-points as carriers.

2 Symmetry

DEFINITION. *A projection* P *is symmetric if*

$$[Pf](x) = [Pf(1-\cdot)](1 - x); \; f \in C[0,1], \; x \in [0,1].$$

The following theorem serves to significantly reduce the number of parameters in both (IIIa') and (IIIb'). In seeking minimal projections, attention need be directed only to those projections having symmetric carriers.

THEOREM. *Amongst the minimal projections from* $C[0,1]$ *to* $P_n[0,1]$ *there exists a symmetric projection.*

Proof. Let $\hat{P}$ be defined by

$$[\hat{P}f](x) = [Pf(1-\cdot)](1 - x); \; f \in C[0,1], \; x \in [0,1].$$

Then $(P + \hat{P})/2$ is a symmetric projection such that

$$\|(P + \hat{P})/2\| \leq \|P\| = \|\hat{P}\|.$$

By restricting attention to symmetric finite carrier projections, problems (IIIa') and (IIIb') may be replaced by:

(IIIa") Minimizing over $[(m - n - 1)(n + 1)/2]$ coefficient parameters.

(IIIb") Minimizing over $[(m - 2)/2]$ carrier parameters.

(The bracket symbol is used here to denote integer part.)

3 Expansion by interpolators

It was shown by Morris and Cheney [2] that any finite carrier projection P may be written in the form

$$P = \Sigma\lambda_k P_k^I, \ \Sigma\lambda_k = 1,$$

where each P_k^I is a projection carried on $n + 1$ points, i.e., an interpolating projection (interpolator). In this representation, the linear parameters of (IIIa") become the coefficients λ_k; and the symmetrization may be accomplished by including in the sum $\hat{P}_k^I$, for each P_k^I, and restricting the coefficients of these interpolators to be the same.

THEOREM. $\|\Sigma\lambda_k P_k^I\|$ is a convex function of $\lambda = (\lambda_1, \ldots)$.

From a computational point of view, this means of representing projections possesses the following advantages:

- for fixed carriers, a convex function is being minimized;
- for fixed carriers, the biorthogonal v's need not be computed for each change in the λ parameters;
- for special cases, the number of λ's and the number of f-values in (II) may be significantly reduced (see the results below for the case of $n = 2$).

4 Numerical results for the quadratic case

For symmetric finite carrier projections from $C[0,1]$ to $P_2[0,1]$, explicit numerical results have been obtained for the cases of 4, 5 and 7 carriers. Two simplifying features were found to be available in these cases, thereby further reducing the complexity of the computations. First, it was reasoned that the worst case, for the maximum over the 2^m f-values of (II), should occur for $f(x) = +1$ $(0 \leq x \leq 1/2)$, $\pm 1 (1/2 < x < 1)$, $-1 (x = 1)$. In each case, the final result was checked for all 2^m of the original choices, and the validity of the preceding observation borne out. As an example, in the case $m = 7$, the number of choices was reduced from 64 (one ± 1 value could be taken to be 1) to 8.

A second observation showed that many of the interpolators would not enter in a minimal projection. This fact was also borne out numerically as well. Those interpolators which did enter were much fewer in number; for example, the reduction was from 10 to 4 interpolators in the case $m = 7$.

In the worst case considered $(m = 7)$, the computations involved:

- two variable carriers
- three linear parameters $(\lambda_1, \lambda_2, \lambda_3)$
- eight choices for $f = \pm 1$
- finding the extrema of a quadratic.

The small number of parameters allowed searches to be carried out rather rapidly on a Hewlett-Packard 9830A programmable desk calculator.

Four Symmetric Carriers (see, also, [3; Theorem 14]).

min norm = 1.24518

carriers: 0, s, 1-s, 1 where s = .46127

Five Symmetric Carriers

min norm = 1.22879

carriers: 0, s, 1/2, 1-s, 1 where s = .2462

$$P = \lambda_1 P_{0,1/2,1} + \frac{\lambda_2}{2}[P_{o,s,1-s} + P_{s,1-s,1}] + \lambda_3 P_{s,1/2,1-s}$$

where $\lambda_1 = .8980$, $\lambda_2 = .0872$, $\lambda_3 = .0148$

Lebesgue function exhibited six extrema.

Seven Symmetric Carriers

min norm = 1.22349

carriers: 0, s, t, 1/2, 1-t, 1-s, 1 where s = .1366,
t = .3938

$$P = \lambda_1 P_{0,1/2,1} + \frac{\lambda_2}{2}[P_{s,t,1-s} + P_{s,1-t,1-s}]$$
$$+ \frac{\lambda_3}{2}[P_{0,s,1-s} + P_{s,1-s,1}] + \frac{\lambda_4}{2}[P_{0,t,1-t} + P_{t,1-t,1}]$$

where $\lambda_1 = .8300$, $\lambda_2 = .1144$, $\lambda_3 = .0319$, $\lambda_4 = .0237$. Lebesgue function exhibited 10 extrema.

In these results, P_{rst} denotes the interpolator carried on r, s, and t. The indicated accuracy reflects the extent to which the authors were able to conduct a semi-automated search using the above-mentioned machine. The authors are indebted to California Investment Counseling for providing the opportunity to "bootleg" time on the H-P 9830A.

References

1 Chalmers, B. L. and F. T. Metcalf, Minimal generalized interpolation projections, submitted.

2 Morris, P. D. and E. W. Cheney, On the existence and characterization of minimal projections, J. für die Reine und Angerwandte Math., 270 (1974), 61-76.

3 Morris, P. D., K. H. Price, and E. W. Cheney, On an approximation operator of de La Vallee Poussin, J. Ayproximation Theory, 13 (1975), 375-391.

B. L. Chalmers F. T. Metcalf
Department of Mathematics
University of California, Riverside
Riverside, California 92502

APPROXIMATION OF FIXED POINTS OF CONTRACTIVE

AND EXPANDING OPERATORS

L. Collatz

This survey discusses some complications which occur in applying approximation methods to problems in sciences, describes some numerical methods which may be helpful in certain cases, and gives numerical results for combination of approximation and iteration methods also in the case of expanding operators.

1 Approximation theory and applications

In applying the theory of approximation of functions to problems of physics, engineering, economics and other sciences, we observe the following complications:

1) One has to use more complicated classes of approximating functions than the usual polynomials, rational functions or exponential functions.

2) Functions of several independent variables occur and one cannot use the theory of Haar systems.

3) Many problems in applications have several solutions, often only one of which is stable.

4) One has to apply chained approximation. This theory is only beginning to be developed.

5) There are often several functions to approximate, and one must consider simultaneous approximation, combi-approximation and field approximation.

6) There occur more complicated restrictions than usually considered. Let us illustrate the situation with some simple examples.

2 Error bounds and chained approximation

The nonlinear boundary value problem for a function $u(x,y)$ using polar coordinates r, ϕ

(2.1) $Tu = -\Delta u - u - e^u = 0$ in $B = \{(x,y),\ r^2 = x^2 + y^2 < 1\}$

(2.2) $u = 0$ on ∂B for $r = 1$

with $\Delta u = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$ may describe the stationary distribution of temperature in a homogeneous circular disc. We look for an approximate solution $w(x,y)$ satisfying the boundary condition

(2.3) $$u \approx w = \sum_{\nu=1}^{p} a\ (1 - r^{2\nu}).$$

Applying one-sided Chebychev-Approximation (Schumaker-Taylor [69]) we wish to choose the parameters a_ν such that

(2.4) $0 \le Tw \le \delta$ in B, $\delta = \text{Min}.$

Then we choose other values $\hat{a}_\nu$ for the a_ν with

(2.5) $-\hat{\delta} \le Tw \le 0$ in B, $\delta = \text{Min}.$

For instance for $p = 2$ we get the approximation problem

$$0 \le Tw = a_1(3 - r^2) + a_2(-1 + 16t^2 + r^4) - \exp[a_1(1 - r^2)$$

$$+ a_2(1 - r^4)] \le \delta \quad \text{for} \quad r \in [0,1], \quad \delta = \text{Min},$$

which is a chained approximation problem because the parameter a_1 (and also a_2) occurs at two different places (for a discussion of chained approximation, see Hoffmann [69, 76], Collatz [75] Collatz-Krabs [73]). Even for $p = 1$

$$(2.6)\quad 0 \le Tw = a_1(3 - r^2) - \exp[a_1(1 - r^2)] \le \delta$$

$$\text{for } r \in [0,1],\quad \delta = \text{Min}$$

we have the complications 1), 3), 4) mentioned in section 1.

Results of the Theory. Fig. 1 gives the graphs of $T[a_1(1 - r^2)]$ for $a_1 = 0.5$ and $a_2 = 0.6191$. The 4 encircled points are an H-set (compare Collatz [56][69], G. D. Taylor [72]). Therefore we have the exact error bounds for the stable solution

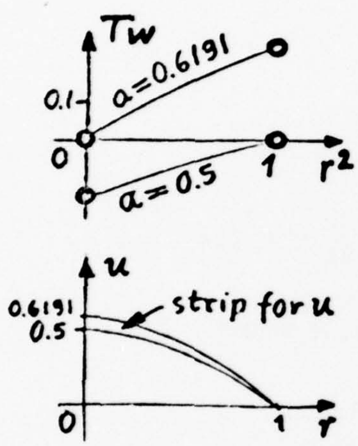


fig.1

$$(2.7)\quad 0.5(1 - r^2) \le u(x,y) \le 0.6191(1 - r^2) \quad \text{in } B,$$

and we know that these are the best possible error bounds one can get in the linear manifold $w = a_1(1 - r^2)$. If one wishes to get sharper error bounds one has to use other manifolds, for instance (2.3) with $p > 1$. Then one can easily improve the bounds (2.7).

3 Chained approximation for nonlinear integral equations

Chained approximation arises very often. We look for an approximate solution $w(x) = e^{ax}$ for the solution $u(x)$ of the Volterra-Equation

$$(3.1)\quad u(x) = Tu(x) = 1 + \int_0^x e^{xt}[u(t)]^2 dt$$

in the interval $J = [0;0.4]$. The Chebychev approximation

problem

$$(3.2)\quad \Phi(a) = \|Tw - w\|_\infty = \operatorname*{Max}_{x\varepsilon J}\left|\frac{e^{x(x+2a)}-1}{x+2a} - e^{ax} + 1\right|, \Phi(a) = \text{Min}$$

is chained and gives the value $a = 1.44$ with $\Phi(a) = 0.060$.

4 Subdomains with error bounds

Let the displacement $u(x,y,t)$ of a membrane be describe by the wave equation

$$(4.1)\quad Lu = \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0$$

and the initial boundary conditions

$$(4.2)\quad u = g(x,y) = (1 - x^2)(1 - y^2),\ \frac{\partial u}{\partial t} = 0 \text{ for}$$

$$(x,y) \varepsilon B = \{|x| < 1,\ |y| < 1\},\ t = 0$$

$$(4.3)\quad u = 0 \text{ for } (x,y) \varepsilon \partial B,\ t \geq 0.$$

We look for an approximate solution $w(x,y,t)$ of the form

$$(4.4)\quad u \approx w(x,y,t) = \sum_{\nu=1}^{p} a_\nu[\cos(b_\nu x)\cos(c_\nu y)$$

$$+ \cos(c_\nu x)\cos(b_\nu y)]\cos((b_\nu^2 + c_\nu^2)^{1/2} t)$$

which satisfies (4.1) and $\partial u/\partial t = 0$ for $t = 0$ for all b_ν, c_ν, and (4.3) for suitable b_ν, c_ν, for instance $b_1 = c_1 = \pi/2$ or $b_2 = \pi/2$, $c_2 = 3\pi/2$. The theorem on pointwise monotonicity applied to a function η with

$$(4.5)\quad L\eta = 0 \quad \text{in } B\times(0,\infty),\ \frac{\partial\eta}{\partial t} = 0 \quad \text{in } B, t = 0,$$

$|\eta| \leq \delta$ in $B, t = 0$ yields the bound $|\eta| \leq \delta$ in

the pyramid P determined by the characteristics (see fig. 2):

$$(4.6)\quad P = \{(x,y,t),\ |x| \leq 1,\ |y| \leq 1,\ t \geq 0,$$

$$\operatorname{Max}(|x|,\ |y|) \leq 1 - t\}.$$

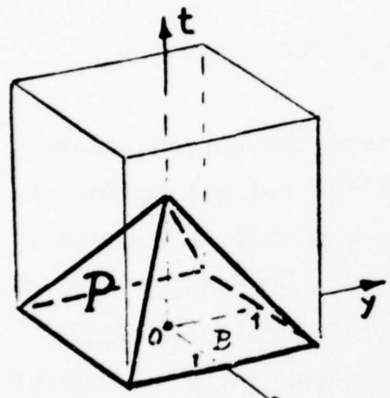


fig 2

We use the Chebychev approximation problem

$$(4.7)\quad \|w(t = 0) - g\|_\infty = \text{Inf}$$

to determine the parameters a . For $p = 2$ we get the values $a_1 = 0.5333$, $a_2 = -0.0394$, and the exact error estimate

$$(4.8)\quad |w - u| \leq 0.0122 \quad \text{in } P.$$

The estimate (4.8) is not assured in the whole domain $B \times (0,\infty)$. This is a typical situation: the approximation methods are numerically useful for getting approximate solutions and in case of validity of monotonicity theorems they also give error bounds.

5 Local relaxation and splines

A function $y(x)$ may be defined in the unbounded interval $J = [0,+\infty)$ by the nonlinear differential equation

$$(5.1)\quad Ty = \frac{dy}{dx} - x^2 + \frac{1}{y} = 0 \quad \text{with } \lim_{x\to+\infty} y(x) = 0.$$

We consider rational Chebychev approximation

$$(5.2)\quad y(x) \approx \tilde{w}(x) = \frac{P(x)}{Q(x)}$$

where P,Q are polynomials with $\partial Q > \partial P$, for instance

$$\tilde{w}(x) = \frac{a + x}{b + ax^2 + x^3}$$

The defect Tw (see Fig. 3) is small for $x \geq x_o$, and can be reduced by local relaxation, changing $\tilde{w}(x)$ into $w(x)$ by

$$(5.3)\quad w(x) = \begin{cases} \tilde{w}(x) \text{ for } x \geq x_o \\ \tilde{w}(x) + \alpha(x - \beta)(x - x_o)^2 \text{ for } 0 \leq x \leq x_o. \end{cases}$$

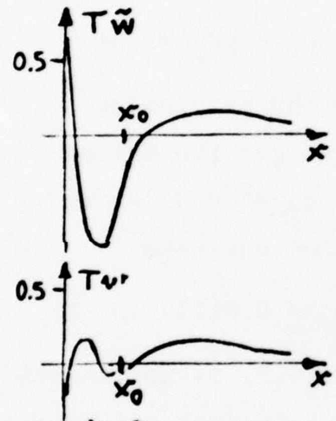


fig. 3

Fig. 3 shows the defect before and after the relaxation for the values $a = 0.4887$, $b = 0.7984$, $x_o = 0.7$, $\alpha = -0.0564$, $\beta = 26.88$. This idea of local relaxation, which was numerically successful also in other problems, is related to the idea of spline-approximation.

6 Several solutions of a problem

Let $w(x,y)$ be an approximate solution for a fixed point $u(x)$ of the Urysohn-equation

$$(6.1)\quad u(x,y) = Tu = -1 + \int_B \frac{2[u(s,t)]^2 dsdt}{2 + x + y + u(s,t)},$$

where B is the square $0 \leq x, y \leq 1$. We choose w in the simplest way as a constant, $x = \text{const.} = a$. This leads to the chained rational Chebychev approximation problem

$$(6.2)\quad \Phi(a) = \|Ta - a\|_\infty = \operatorname*{Max}_B \left|-1 + \frac{2a^2}{2 + x + y + a} - a\right|,$$

$$\Phi(a) = \text{Min}$$

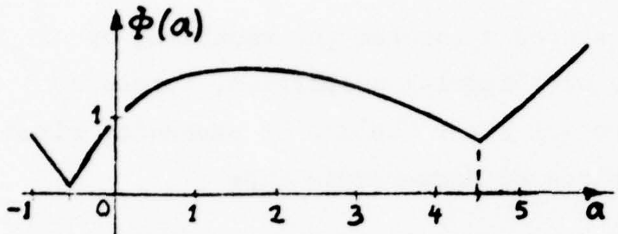


fig. 4

Fig. 4 shows the graph of $\Phi(a)$ with an absolute minimum $a = -0.615$ near the stable solution and a local minimum $a = 4.503$ near the unstable solution with expanding behaviour.

7 Numerical procedure for expanding operators

The following iteration procedure has been tested numerically for constractive as well as for expanding operators T. We seek a solution u of $u = Tu$. Let $v_o(x_j, a_1, \ldots, a_p)$ and $v_1(x_j, a_1, \ldots, a_p)$ be a pair of functions of $x_1, \ldots, x_n$ with

$$(7.1)\quad v_1(x_j, a_1, \ldots, a_p) = T\, v_o(x_j, a_1, \ldots, a_p).$$

We describe the iteration by induction. If we have determined in the $(p - 1)$ step values $a_s^{(p-1)}$ for the $(p - 1)$ parameters a_s $(s = 1, \ldots, p-1)$, then we calculate values $a_s^{(p)}$ for the p parameters a_s $(s = 1, \ldots, p)$ by

$$(7.2)\quad \Phi(a_1, \ldots, a_p) = \|v_1 - v_o\|_\infty;\ \phi(a_1, \ldots, a_p) = \text{Inf}.$$

To get the infimum of ϕ approximately, one can start by

using for the a_s $(s = 1,\ldots,p-1)$ the values $a_s^{(p-1)}$ one has got in the $(p - 1)$ step. Thus, at first one takes only one parameter and then enlarges the number of parameters with every step.

8 Expanding operators and nonlinear integral equations

Krasnoselskii [64] stated a theorem for expanding operators. Based on cones with special properties, Sprekels [75] used this theorem to get error bounds for expanding fixed points of integral equations of Hammerstein type

$$(8.1) \quad u(x) = Tu = \int_B K(x,s)f(s,u(s))ds$$

with positive kernel $K(x,s)$. He could get good error bounds if one has a good approximation of $K(x,s)$ as a product $p(x)\cdot q(s)$. This leads to the following approximation problem:

$$(8.2) \quad A \leq \frac{K(x,s)}{p(x)q(s)} \leq A\cdot\sigma$$

with $A = \text{const}$, $\sigma = \text{const}$, $\sigma = \text{Min}$. Sprekels [75] applied his theory to nonlinear differential equations. I thank him for the following example:

$$(8.3) \quad -y''(x) = \lambda[y(x)]^4, \quad 0.2y(0) - y'(0) = -0.1y(1) + y'(1) = 0$$

He got the following inclusion, valid for every $\lambda > 0$:

$$\lambda^{-1/3}\, 0.4174\, \frac{9 + x}{12} \leq y(x) \leq \lambda^{-1/3}\, 0.5846.$$

9 Secant method (Regula falsi) and Newton-iteration

Suppose that in the interval $J = [a,b]$ the real-valued function $f(x) \in C^2(J)$ has fixed signs of $f'(x)$ and $f''(x)$; and that $f(x)$ has a zero $z \in (a,b)$ with $f(z) = 0$. We start with two values $x_o, x_1 \in J$ and calculate x_2 with the secant method:

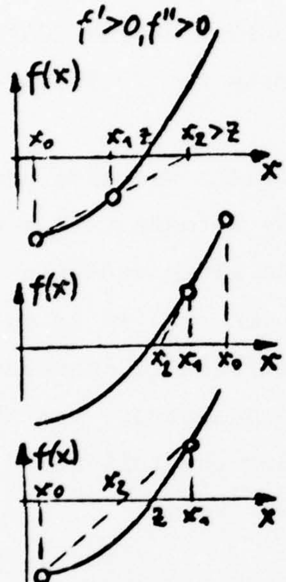


fig.5.

$$(9.1) \quad x_2 = x_1 - \frac{x_1 - x_o}{f_1 - f_o} f_1 \text{ with}$$

$$f_j = f(x_j) \ (j = 0,1).$$

We suppose $x_2 \in J$. Then one gets lower or upper bounds for z by choosing x_o, x_1 on the same or opposite side of z, respectively, as shown in the following table:

	$f'(x)$ and $f''(x)$ same sign	$f'(x)$ and $f''(x)$ opposite sign
x_o, x_1 on the same size of z	$x_2 > z$	$x_2 < z$
x_o, x_1 on opposite sides of z	$x_2 < z$	$x_2 > z$

W. Hofmann [71] gives a generalization to equations $Tu = \theta$ in Banach spaces with θ the zero-element. He is working with linear "difference-operators" $T[u',u'']$ with the property $T[u',u''](u' - u'') = Tu' - Tu''$ which are described in detail in his paper. Under certain conditions he can prove inclusion theorems similar to the theorems given in the table. If T is Fréchet-differentiable, one gets similar inclusion

theorems also for the Newton procedure. Good starting elements x_o, x_1 are essential; one gets these elements with the aid of Approximation Theory. I thank Mr. Hofmann for the numerical example: For $Tu = -\Delta u + 2u^2 = 0$ in $B: (r^2 = x^2 + y^2 < 1)$ and $u = 1$ on $\partial B: (r = 1)$ monotonicity and Schauder's Fixed Point Theorem (compare Collatz [66]) give the bounds (easy to get) $1 - (1/2)(1 - r^2) \leq u \leq 1 - (1/4)(1 - r^2)$ with $|0.625 - u(0,0)| \leq 0.125$ (error 20%). The method of Hofmann improves this estimate to $|0.70594 - u(0,0)| \leq 0.00484$ (error 0.68%).

The idea of using monotonicity properties was also used by Voss [76], but in a different way as by Hofmann. Voss uses a decomposition of the operator T in the given equation $Tu = \theta$; he writes $T = T_1 - T_2$ and supposes T_1 is an operator with the same properties as the "difference operator" of Hofmann mentioned above, while T_2 is a monotone operator. To illustrate the decomposition we consider the following initial-value problem for which I thank Dr. Voss:

$$y'' - y^2 + 2y = 2x^2 = 0, \ y(0) = 1, \ y'(0) = 0$$

Here a possible decomposition is

$$T_1 y = (y'' - y^2, y(0), y'(0)) \text{ and } T_2 y = (-2y + 2x^2, \ 1, 0)$$

and $Ty = T_1 y - T_2 y$ maps the solution into the zero element of the space $C^2[0,1/2] \times R^1 \times R^1$.

Taking quadratic polynomials, one gets the error bounds $1 - (1/2)x^2 \leq y(x) \leq 1$.

10 Inverse problems and not well posed problems

We consider the case that in the heat-conduction equation (or equation for diffusion, concentration a.o.)

$$\text{(10.1)} \quad Lu = \frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left[k(x)\,\frac{\partial u}{\partial x}\right] = 0$$

the unknown coefficient $k(x)$ should be determined from the observations of the temperature $u(x,t)$ of a beam at two different times. As a simple case suppose

(10.2) $u = 0$ for $x = \pm 1$, $0 \le t \le 1$ (at the ends of the beam

$$u = 1 - x^2 \text{ for } |x| \le 1,\ t = 0$$

$$u = \frac{1}{2}(1 - x^2) \text{ for } |x| \le 1,\ t = 1.$$

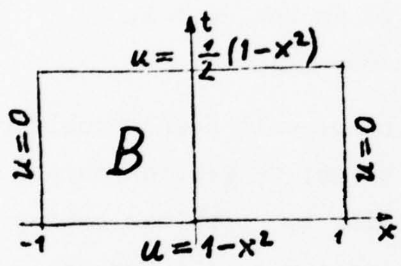


fig. 6

We take as approximate terms

$$\text{(10.3)} \quad u \approx w(x,t) = (1 - x^2)\left[1 - \frac{1}{2}t + (t - t^2)\cdot\sum_{\nu=0}^{p} a_\nu t^\nu\right]$$

$$\text{(10.4)} \quad k(x) = \sum_{\mu=0}^{q} b_\mu x^{2\mu}$$

The coefficients a_ν, b_μ are calculated from the Chebychev Approximation

$$\text{(10.5)} \quad \|Lw\|_\infty \overset{!}{=} \text{Inf.}$$

with the maximum norm in the domain $B: |x| \le 1,\ 0 \le t \le 1$ (see Fig. 6). One gets the values

3 Parameters	5 Parameters
$a_o = -0.1716$	$a_o = -0.19240$ $a_1 = 0.03986$
$b_o = 0.3431$ $b_1 = -0.1122$	$b_o = 0.34662$ $b_1 = -0.11571$ $b_2 = 0.00001$
$\|\|Lw\|\|_\infty = 0.0147$	$\|\|Lw\|\|_\infty = 0.00088$

The function Lw has 5 extrema in $B^+ = B \cap \{x \geq 0\}$; the minimum of the modulus of these 5 extrema is 0.00083. Fig. 7 shows the graph of $k(x)$ for $q = 2$, the domain B with the dotted line $Lw = 0$, and the extremal points of Lw. It is remarkable that for $k(x)$ one gets nearly the same result for $q = 2$ as for $q = 1$, (difference $\approx$ 1%).

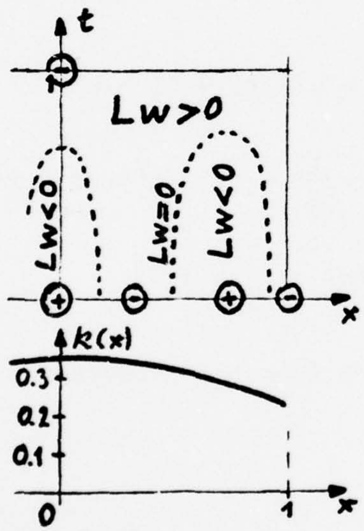


fig.7

Many other not well posed problems are well known; we mention only the following two problems coming from applications: 1.) For the wave equation with given coefficient $k(x)$

$$(10.6) \quad Lu = \frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x}\left(k(x)\frac{\partial u}{\partial x}\right) = 0$$

we consider the observed displacement $u(x,t)$ for $|x| \leq 1$, $t = 0$ and $t = 1$ (as in Fig. 6). For $x = \pm 1$, $0 \leq t \leq 1$ we suppose $u = 0$, and we want to calculate $u(x,t)$ in the interior of the rectangle B.

2.) In the equation for diffusion $u(x,t)$

$$(10.7) \quad Lu = \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 ,$$

we wish to calculate $u(x,t)$ for $x = x_o$, $t > 0$, if u is given for $x = 0$, $t > 0$ and $0 < x < x_o$, $t = 0$ and $\partial u/\partial x$ is given for $x = 0$, $t > 0$ (see Fig. 8). Here we have to apply simultaneous approximation, compare Bredendiek [69], [70], [76].

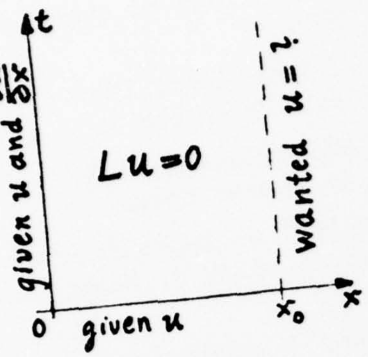


fig. 8

Of course, it is easy to enlarge the number of examples for which the numerical methods of approximation theory are useful.

The author thanks Mrs. Susanne Böttger and Mr. Rolf Wildhack very much for numerical calculation on a computer.

References

1 Bredendiek E. [69] Simultan Approximation, Arch. Rat. Mech. Anal., 33 (1969), 307-330.

2 Bredendiek, E. [70] Charakterisierung und Eindeutigkeit bei Simultanapproximationen. Z. Angew. Math. Mech. 50 (1970), 403-410.

3 Bredendiek, E. and L. Collatz [76] Simultan Approximation bei Randwertaufgaben, to appear in Intern. Ser. Num. Math. 1976.

4 Collatz, L. [56] Approximation von Funktionen bei einer und bei mehreren unabhängigen Veränderlichen. Z. Angew. Math. Mech. 36 (1956), 198-211.

5 Collatz, L. [69] Nichtilineare Approximationen bei Randwertaufgaben, V. IKM. Weimar, (1969), 169-182.

6 Collatz, L. [75] Bemerkungen zur verketteten Approximation, Intern. Ser. Num. Math. 26 (1975), 41-45.

7 Collatz, L. [66], Functional Analysis and Numerical analysis, Acad. Press 1966, 473 p.

8 Collatz, L. und W. Krabs [73], Approximationstheorie Teubner, Stuttgart, 1973, 208 S.

9 Hofmann, W. [71] Monotoniesätze fur Regula falsi und Newton-Verfahren, Computing 8 (1971), 143-156.

10 Hoffmann, K. H. [69] Zur Theorie der nichtlinearen Tschebyscheff Approximation mit Nebenbedingungen Numer. Math. 14 (1969), 24-41.

11 Hoffmann, K. H. [76] Über verkettete Approximation, to appear in Intern. Ser. Numer. Math., 1976.

12 Krasnoselskii, M. A. [64] Positive solutions of operator equations. Nordhoff Groningen 1964.

13 Meinardus, G. [67] Approximation of Functions, Theory and Numerical Methods, Springer Verlag, 1967, 198 S.

14 Sprekels, J. [75] Existenz und 1. Einschließung bei superlinearen Randwertaufgaben, to appear soon.

15 Schumaker, L. L. and Taylor, G. D. [69] On Approximation by Polynomials having restricted ranges, SIAM J. Numer. Anal. 6 (1969), 31-36.

16 Taylor, G. D. [72] On minimal H-sets, J. Approximation Theory 5 (1972), 113-117.

17 Voss, H. [76], Ein neues Verfahren zur Einschließung der Lösungen von Operatorgleichungen, to appear in ZAMM.

Lothar Collatz
Institut für Angewandte Mathematik
Universität Hamburg
Bundesstr. 55
2000 Hamburg 13, Germany

ON SATURATION THEOREMS

Z. Ditzian

For an approximation process $A_n f$ a saturation theorem determines the optimal rate of convergence. This rate, being exceeded only for a limited, "trivial", class of functions, is achieved on a "large" class of functions. Saturation classes and theorems were introduced by Favard and investigated by many authors for various operators, but as far as the author is aware, the "trivial" class was always a finite dimensional space at most. The main difference here is that we will obtain an infinite dimensional but non-dense subspace for our trivial class.

We shall deal with convolution approximation operators defined as follows:

$$(1.1)\quad A_n f(x) \equiv \int_{R^k} f(x-t)\, d\mu_n(t)$$

where $t = (t_1,\cdots,t_k)$, $x = (x_1,\cdots,x_k)$ and

$$(1.2)\quad \int_{R^k} d\mu_n(t) = 1 \quad \text{and} \quad d\mu_n(t) \geq 0$$

and for some sequence $\sigma_n^2 = o(1)$, we have

$$(1.3)\quad \lim_{n\to\infty} \sigma_n^{-2} \int_{R^k} t_i t_j\, d\mu_n(t) = A_{ij}, \quad \det (A_{ij}) \neq 0,$$

$$(1.4)\quad \lim_{n\to\infty} \sigma_n^{-2} \int_{R^k} t_i\, d\mu_n(t) = 0 \quad 1 \leq i \leq k$$

and

$$(1.5)\quad \lim_{n\to\infty} \sigma_n^{-2} \int_{R^k - B(0,\delta)} d\mu_n(t) = 0 \quad \delta > 0 \quad B(0,\delta)=\{t;|t|<\delta\}.$$

One can check that many convolution approximation operators satisfy the above conditions. For instance:

(a) $$A_n f(x) = \gamma_n \int_{R^k} \exp\{-n \sum_{i=1}^{k} t_i^2\} f(x-t)dt;$$

(b) $$A_n f(x) = \gamma_n(r) \int_{R^k} \prod_{i=1}^{k} \left(\frac{\sin(n\, t_i/2)}{\sin(t_i/2)}\right)^{2r_i} f(x-t)dt,$$

r_i integer, $r_i \geq 2$;

(c) $$A_n f(x) = [V(B(0,\tfrac{1}{n}))]^{-1} \int_{B(0,\frac{1}{n})} f(x-t)dt;$$

or

(d) $$A_n f(x) = \gamma_n \int_{E(0,\frac{1}{n})} f(x-t)dt,$$

where

$$E(0,\tfrac{1}{n}) = \{t:\ (t_1+t_2)^2 + \frac{(t_1-t_2)^2}{2} + \frac{1}{3}t_3^2 + \cdots + \frac{1}{k}t_k^2 = \frac{1}{n^2}\}.$$

2. The local small o saturation result

The key to our investigation is the result of this section. (However this is not the most difficult result.)

THEOREM 2.1. *For* $A_n f$ *satisfying* (1.1), (1.2), (1.3), (1.4) *and* (1.5), *we have*, *for* $E(B_r)$ *being* $L_p(B_r)$ $1 \leq p < \infty$ *or* $C(B_r)$ (where $B_r \equiv B(x_o,r)$),

(a) $\|A_n f-f\|_{E(B_r)} = o(\sigma_n^2)$ *implies* $P(D)f = \sum_{j,i} A_{ij} \frac{\partial^2 f}{\partial t_i \partial t_j} = 0$ *in* B_r *and* $f \in C^\infty$ *in* B_r;

(b) $P(D)f = 0$ *in* B_r *implies* $\|A_n f-f\|_{E(B_{r-\delta})} = o(\sigma_n^2)$.

Proof. To prove (a), we observe that for $\phi \in C_o^\infty(B_r)$

(supp $\phi \subset B_r$) we have

$$\lim_{n\to\infty} \langle \sigma_n^{-2}(A_n f - f), \phi\rangle = 0,$$

since $\phi \in L_p^*$ $1 \leq p < \infty$ and $\phi \in L_1$ and therefore $\int_A \phi dx = \mu(A)$ is a measure and in the dual of C. This implies

$$\langle \sigma_n^{-2}(A_n f - f), \phi\rangle = \langle f, \sigma_n^{-2}(A_n^* \phi - \phi)\rangle =$$

$$\langle f, P(-D)\phi\rangle + o(1) = \langle f, P(D)\phi\rangle + o(1) = o(1)$$

or $P(D)f = 0$ weakly as a distribution in B_r. Since $P(D)$ is elliptic (as can easily be shown), Apriory estimates [1, p.66] can be used to show that $f \in C^\infty(B_r)$, and therefore $P(D)f = 0$ not only in the distributional sense.

The proof of (b) is mere computation if we know that $P(D)f = 0$ in $E(B_r)$ implies $f \in C^\infty(B_r)$. This can be obtained by using some Apriory estimates.

3. The local big 0 theorem

THEOREM 3.1. *Let $A_n f$ be defined as in Theorem 2.1, then for $1 < p < \infty$ we have*

(a) $\|A_n f - f\|_{L_p(B_r)} = 0(\sigma_n^2)$ *implies* $f \in W_{2,p}(B_{r-\delta})$;

(b) $f \in W_{2,p}(B_r)$ *implies* $\|A_n f - f\|_{L_p(B_{r-\delta})} = 0(\sigma_n^2)$

(*where* $W_{2,p}$ *is the Sobolev space of functions with* 2 *derivatives in* L_p).

THEOREM 3.2. *Let $A_n f$ be defined as in Theorem 2.1, then for* $E = C$ (*or* $E = L_1$)

(a) $\|A_n f - f\|_{E(B_r)} = 0(\sigma_n^2)$ *implies* $P(D)f \in L_\infty(B_{r-\delta})$

(or $\int_e P(D)f\ dx = \mu(e)$ <u>is a measure of bounded variation on</u> $B_{r-\delta}$) <u>where</u> $P(D)f$ <u>is taken in the distributional sense</u>;

(b) $f \in L_{2,\infty}(B_r)$ <u>implies</u> $\|A_n f-f\|_{C(B_{r-\delta})} = O(\sigma_n^2)$;

$f \in L_{1,1}(B_r)$ <u>and the derivative of</u>

$\frac{\partial}{\partial\xi} f$ <u>is a measure of bounded variation implying</u>

$\|A_n f-f\|_{L_1(B_{r-\delta})} = O(\sigma_n^2)$.

It is not possible to give a complete proof in the small space allotted to this paper. However we note that the main difference between 3.1 and 3.2 is a result of the developed a priori estimate for L_p, $1 < p < \infty$ (see Nirenberg [6] and Calderon Zygmund [3]), which does not apply to L_1 nor L_∞. Other parts of the proof follow in part the author's result for convolution operators [4] as well as those of Kozima and Sunouchi [5] and Butzer Nessel [2].

4. Global results

Many of the above results have global analogues. For instance, in the case of the Gauss-Weierstrass transform:

(4.1) $G_n * f = \gamma_n \int \exp\left(-n(x-y)^2\right) f(y)dy \quad x,y \in R^k$,

it was proved that $\|G_n*f-f\|_{L_p} = o\left(\frac{1}{n}\right)$ if and only if $f = 0$. It is clear that $P(D) = \Sigma\ \partial^2/\partial x_i^{\ 2}$ and indeed the only harmonic function that belongs to L_p is $f = 0$. However $\|e^{-x^2}(G_n*f-f)\|_E = o\left(\frac{1}{n}\right)$, $n \geq 2$, implies f is harmonic but $|f| = o(e^{|x|^2})$ $x \to \infty$. For other sequences of operators different weight functions may be needed.

5. Remarks

The results are applicable to convolution operators on

π^k where the functions are periodic. A non-convolution multidimensional sequence of positive operators leads sometimes (as in the case of Bernstein or Baskakov operators) to a solution of elliptic equations with non-constant coefficients.

References

1 Agmon, S., Lectures on elliptic boundary value problems, Van Nostrand, 1965.

2 Butzer, P.L. and R.J. Nessel, Contribution to the theory of saturation for singular integrals in several variables, Indag. Math., 28 (1966), 515-531.

3 Calderon, A.P. and A. Zygmund, On the existence of certain singular integrals, Acta Math., 88 (1952), 85-139.

4 Ditzian, Z., Saturation classes for a sequence of convolution operators with limited oscillation, Israel Jour. of Mathematics, 17 (1974), 315-324.

5 Kozima, M. and G.I. Sunouchi, On the approximation by a general singular integral, Tôhoku Math. Jour., 20 (1968), 146-169.

6 Nirenberg, L., Estimate and existence of solutions of elliptic equations, Comm. Pure Appl. Math., 9 (1956), 509-530.

Z. Ditzian
Department of Mathematics
University of Alberta
Edmonton, Alberta T6G 2G1
Canada

EXTREMAL PROBLEMS ON POLYNOMIALS

P. Erdős

The subject has very many aspects and ramifications and I am not at all competent to deal with many of them and in fact in this short article I do not even have the space. Thus I will eventually concentrate on problems which my collaborators and I investigated. I will try to review them as systematically as possible and will also state a few new problems.

Herzog, Piranian and I in a paper [7] (we will refer to this paper as (I)) investigated several geometric extremal problems for polynomials; we stated there many conjectures some of which were settled (positively or negatively) by Pommerenke and others. First I give a short review of the fate of the conjecture stated in (I). Pommerenke's two papers in which he deals with the problems in (I) are Pommerenke [12] and [14]; we will refer to them as (PI) and (PII), respectively. Some of these problems are discussed by Pommerenke also in [13].

1. Let $-1 \leq x_1 \leq \dots \leq x_n \leq 1$. We conjectured and Elbert [4] proved that the measure of the set of real numbers x for which $|\Pi_{i=1}^{n}(x - x_i)| \leq 1$ is not greater than $2\sqrt{2}$. The polynomial $1 - x^2$ shows that this is best possible. The proof of Elbert is surprisingly complicated and a simpler one would be desirable.

2. Put $f_n(z) = \Pi_{i=1}^{n}(z - z_i)$. Denote by $E_n(f)$ the set of points where $|f_n(z)| \leq 1$. Pólya proved that the area $A_n(E)$ of $E_n(f)$ is $\leq \pi$, with equality only for z^n and he also proved that the projection of $E_n(f)$ on every line has

Preceding page blank

measure < 4. Denote $\varepsilon_n = \min A_n(E)$ where the minimum is to be taken over all $f_n(z)$ where roots are all in $|z| \leq 1$. Using a deep theorem of G. MacLane we proved $\varepsilon_n \to 0$. We conjectured that $\varepsilon_n > n^{-\eta}$ for every $\eta > 0$ and $n > n_o(\eta)$, we also conjectured that one can place a circle of radius Cn^{-1} is $A_n(E)$. Both these conjectures are still open, but Pommerenke proved some weaker results. More generally in (I) we conjectured that if T is any set of transfinite diameter 1 then to every $\varepsilon > 0$, for $n > n_o(\varepsilon)$ there is a polynomial $f_n(z) = \Pi_{i=1}^n (z - z_i)$ with $z_i \in T$, so that $A_n(E) < \varepsilon$. This is easy if T is $(-2,+2)$ but the general case is open.

Netaniahu and I [8] proved that there is a constant ρ_c depending on c alone so that if T is a connected set of transfinite diameter $1-c$ then $E_n(f)$ contains a circle of radius ρ_c. Our proof is a pure existence proof and does not give an explicit bound for ρ_c. It would be interesting to determine (or estimate) the best possible value of ρ_c and also the best possible value of $A_n(E)$. The following special case might be of some interest. Let the roots of $f_n(z)$ be in $|z| \leq 1 - c$. Then $E_n(f)$ contains the circle $|z| \leq c$ and thus has area $\geq \pi c^2$. I wonder if this is best possible: Is it true that to every $\varepsilon > 0$ there is an $f_n(z)$ with $|z_i| \leq 1 - c$ and $A_n(E) < \pi c^2 + \varepsilon$?

We proved in (I) that there is a polynomial $f_n(z)$, $|z_i| \leq 1$, for which $E_n(f)$ has $n - 1$ components and that it never can have n components. As far as we know the following question has not been investigated: Let T be a set of transfinite diameter 1. Is it true that for each n there is a polynomial $f_n(z)$ all whose roots are in T and for which $E(f)$ has $\geq n - \gamma_n$ components where γ_n is "small" (certainly $o(n)$ and perhaps bounded; if T is $(-2,+2)$, then $\gamma_n = 0$)? If T has transfinite diameter $1 - c$

then in general $\gamma_n > f(c)n$, but as far as we know also this question has not been investigated.

3. A pretty theorem of H. Cartan states that for a polynomial $f_n(z) = \Pi_1^n(z - z_i)$, $i = 1,\dots,n$, the lemniscate $E_n(f)$ can always be covered by circles the sum of whose radii is less than $2e$. We conjectured in (I) that the correct value here should be 2 (clearly best possible if true). We also conjectured that if $E_n(f)$ is connected then it is contained in a circle of radius 2 and center $\frac{1}{n}\Sigma_{i=1}^n z_i$. This conjecture was proved by Pommerenke (PI).

Two conjectures of (I) which seem to me the most attractive are as follows: The sum of the diameters of the components of $E_n(f)$ is $\leq n2^{1/n}$ (equality for $z^n - 1$) and the length of the curve $|f_n(z)| = 1$ is maximal for $f_n(z) = z^n - 1$. As far as we know no real progress has been made with these conjectures though Pommerenke (PII) proved that the length of $|f_n(z)| = 1$ is less than $74n^2$--whereas the "truth" should be $2n + O(1)$.

In (I) we made the ill fated conjecture that the number of components of $E_n(f)$ which have a diameter $> 1 + c$ is less than δ_c, δ_c bounded. Pommerenke (PII) showed that nothing could be farther from the truth, in fact he showed that to every ε and k there is an $E_n(f)$ which has more than k components of diameter $> 4 - \varepsilon$. Our conjecture can probably be saved as follows: Denote by $\phi_n(c)$ the largest number of components of diameter $> 1 + c$ which $E_n(f)$ can have. Surely for every $c > 0$, $\phi_n(c) = o(n)$ and hopefully $\phi_n(c) = o(n^\varepsilon)$. I have no guess about a lower bound for $\phi_n(c)$ also I am not sure whether the growth of $\phi_n(c)$, $1 < c < 4$, depends on c very much.

In (I) we conjectured that if $E_n(f)$ is connected its circumference is $\geq 2\pi$, equality only for z^n. In (PI)

Pommeneke proved this. As far as I know the following question has not yet been investigated. Assume that $E_n(f)$ is convex; how large can its circumference be? Perhaps 8 is the correct bound.

Assume that $E_n(f)$ is connected, let d be its diameter, let b be its width. We conjectured that $b \leq 2$ but in (PII) Pommerenke showed that $b > 2.386$ is possible, he further showed that $b < 2.920$ and that to every $E_n(f)$ there is a direction so that the projection of $E_n(f)$ on this direction has measure < 3.30.

We conjectured that if $E_n(f)$ is connected then to every z_o on the boundary of $E_n(f)$ there is a $z \in E_n(f)$ with $|z - z_o| \geq 2$. Pommerenke in (PII) disproves this and shows that 2 can be replaced by $(3/4)\sqrt{3}$ which is best possible.

Let $|z_i - z_j| \leq 2$, $1 \leq i < j \leq n$. We conjectured that $\Pi_{1\leq i<j\leq n}|z_i - z_j|$ assumes its maximum if the z_i are the vertices of a regular polygon of diameter 1. Danzer and Pommerenke [3] disproved this conjecture for even n but it probably holds for odd n. As far as I know it is open for $n \geq 5$.

Assume that $E_n(f)$ has n components. Is it true that its area is maximal if $f_n(z) = z^n - 1$? If true this probably will not be easy to prove. On the other hand the following conjecture (if true) would be probably not difficult to establish. Let $|z_i| = 1$, $1 \leq i \leq n$. Then there is a path joining the origin to the circle $|z| = 2^{1/n}$ on which $|\Pi_{i=1}^n (z - z_i)| \leq 1$.

4. Many of these problems can be extended to higher dimensional spaces or metric spaces. Let us here restrict ourselves to three dimensional Euclidean space with distance $d(z,z')$. If $z_1,\ldots,z_n$ are n points in space, denote by

$E_n(z_i)$ the set of z's for which $\Pi_{i-1}^n d(z,z_i) \leq 1$. Is it true that if $E_n(z_i)$ is connected it is in the interior of the sphere of radius 2 and center $\frac{1}{n}\Sigma_{i=1}^n z_i$? What is the maximum of the volume of $E_n(z_i)^2$? I first thought that the maximum will be attained for the unit sphere, i.e., in case when all the points z_i coincide, but Piranian showed that this is false already for $n = 2$. It would be very interesting to determine the maximal volume of $E_n(z_i)$ and the distribution of the z_i which achieves this maximum.

Many of these problems are of great geometric interest. In the plane the regular polygon usually gives the solution, but nothing corresponds to this in three dimensions. Here is a typical example: A well known theorem of Pólya states that if $|z_i| \leq 1$ then $|\Pi_{1\leq i<j\leq n}(z_i - z_j)|$ is maximal if the z_i are the vertices of the regular n-gon. Assume now that the z_i are in the unit sphere; I can not even guess for which set of n points z_i, $1 \leq i \leq n$ is the maximum of $|\Pi_{1\leq i<j\leq n} d(z,z_i)|$ assumed. Similarly I can not guess for which set z_i is the area of $\Pi_{1\leq i<j\leq n} d(z_i,z_j) = 1$ maximal? (the z_i are unrestricted here).

5. Let $f_n(z) = \Pi_{i=1}^n(z - z_i)$ and assume that $E(f)$ is connected. I conjectured that $\max_{z\varepsilon E_n(f)} |f_n'(z)| < n^2/2$. Pommerenke proved this with $en^2/2$ instead of $n^2/2$. Let α_i be the distance of z_i to the boundaries of the lemniscate. Perhaps $\Sigma_{i=1}^n \alpha_i \geq n(2^{1/n} - 1)$; equality for $z^n - 1$. $\Sigma_{i=1}^n \alpha_i > c$ could remain true if instead of connectedness we assume $|z_i| \leq 1$.

As far as I remember we never considered the following questions which are perhaps not uninteresting. Let $E_n(f)$ be connected, when is its area minimal? Probably when the roots are all real and $\max|f(z)| = 1$ between any two consecutive roots. A related question: Let $E_n(f)$ be connected.

When is $\int_{E_n(f)} |f_n(z)|$ (area integral) maximal and minimal? The maximum is probably achieved for z^n and the minimum when the roots are all real and between any two roots $|f_n(z)|$ assumes the value 1.

In some cases connectedness of $E_n(f)$ can perhaps be replaced by the following condition: No line separates the components of $E_n(f)$. For example: Is it true that $E_n(f)$ is contained in a circle of radius 2 and center $\frac{1}{n}\Sigma_{i=1}^n z_i$ under this assumption? For this section compare [9].

6. I conjectured 35 years ago that if $f_n(\theta)$ is a trigonometric polynomial whose maximum is 1 and all whose roots are real then

$$\int_0^{2\pi} |f_n(\theta)| \leq 4.$$

Let the rational polynomial $f_n(x)$ have all its roots in $(-1,+1)$, $\max_{-1\leq x\leq 1} |f_n(x)| = 1$ and let x_i, x_{i+1} be two consecutive roots of $f_n(x)$. Then

$$\int_{x_i}^{x_{i+1}} |f_n(x)| \leq d_n(x_{i+1} - x_i)$$

where

$$\frac{1}{y_{i+1} - y_i} \int_{y_i}^{y_{i+1}} T_n(y) = d_n,$$

T_n is the Tchebicheff polynomial and y_i, y_{i+1} are two consecutive roots of T_n.

These conjectures and more have all been proved recently by Kristiansen [11] and Saff and Sheil-Small [15].

I proved [5] that the arc length from 0 to 2π of a trigonometric polynomial f_n of degree n satisfying $|f_n(\theta)| \leq 1$ is maximal for $\cos n\theta$. Let $0 < a < b < 2\pi$,

is it still true that the variation and arc-length in (a,b) is maximal for $\cos(n\theta + \alpha)$ for a suitable α? Also if $|f_n(x)| \leq 1$, $-1 \leq x \leq 1$ is a rational polynomial of degree n, is it true that $T_n(x)$ has the greatest arc-length? The answer clearly must be yes--only a proof is needed.

7. I would like to call attention to two old problems of mine which perhaps belong more to number theory. Let z_n, $n = 1, 2, \ldots, |z_n| = 1$ be an infinite sequence. Put

$$A_n = \max_{|z|=1} \prod_{i=1}^{n} |z - z_i|.$$

Prove (or disprove) $\overline{\lim}\, A_n = \infty$. This problem is probably difficult. Let B_k be the least upper bound of the numbers

$$\left|\sum_{i=1}^{m} z_i^k\right|, \; m = 1,2,\ldots.$$

It is easy to see that a sequence $\{z_j\}$, $j = 1,2,\ldots$ exist for which $B_k < Ck$. I conjecture $\overline{\lim}(B_k/k) > 0$. Clunie [2] proved $\overline{\lim}(B_k/k^{1/2}) > 0$.

These two problems really belong to a chapter called irregularities of distribution of diaphantine analysis, a subject to which K. F. Roth contributed many deep results.

8. Finally I state a few miscellaneous problems on polynomials. First an old problem of mine: Let $-1 \leq x_1 < \ldots < x_n \leq 1$. Let $\ell_k(x_k) = 1$, $\ell_k(x_i) = 0$ for $k \neq i$ be the fundamental functions of the Lagrange interpolation. Prove (or disprove) $(x_0 = -1,\ x_{n+1} = 1)$ that

$$(1) \qquad \min_{0 \leq i \leq n+1} \; \max_{x_i < x < x_{i+1}} \; \sum_{k=1}^{n} |\ell_k(x)| < c \log n.$$

I proved the much weaker result with $n^{1/2}$ instead of $c \log n$. Perhaps I overlooked a simple approach but I

never got anywhere with (1).

The following questions can not be difficult and the answers are perhaps known: Let $z_1,\dots,z_n$ be n points on the unit circle. Let $P_n(z) = \Pi_{i=1}^n (z - z_i)$, let y_i be the point on the arc (z_i, z_{i+1}) where $|P_n(z)|$ assumes its maximum. Is it true that

$$(2) \qquad \prod_{i=1}^{n} |P_n(y_i)| \le 2^n \text{ and } \sum_{i=1}^{n} |P_n(y_i)| \ge 2n?$$

There is equality in (2) for $z^n - 1$.

Let $|w_j| = 1$, $1 \le j \le n+1$, $P_n(z) = \Pi_{i=1}^n (z - z_i)$, $|z_i| = 1$. But

$$A_n(w_1,\dots,w_{n+1}) = \min_{P_n} \max_{1 \le j \le n+1} |P_n(w_j)|.$$

Is it true that $A_n(w_1,\dots,w_{n+1})$ is maximal if $w_j^{n+1} = 1$? i.e., if the w_j's are the (n+1)-st roots of unity. Determine the extreme value. This surely must be simple but at the moment I do not know the answer.

Let $|z_i| \le 1$, $1 \le i \le n$, $f_n(z) = \Pi_{i=1}^n (z - z_i)$. Put

$$A(f_n) = \max_{0 \le r \le 1} \max_{|z|=r} |f_n(z)| \min_{|z|=r} |f_n(z)|.$$

How large is $\max_{f_n} A(f_n)$?

Some of these questions may not be "serious" Mathematics but I am sure the following final problem considered by D. J. Newman and myself for a long time is both difficult and interesting: Let $\varepsilon_k = \pm 1$. Is it true that there is an absolute constant c so that for every choice of the ε_k's

$$\max_{|z|=1} \left| \sum_{k=1}^{n} \varepsilon_k z^k \right| > (1 + c) n^{1/2}?$$

This probably remains true if the condition $\varepsilon_k = \pm 1$ is replaced by $|\varepsilon_k| = 1$. For this section, see [1], [6].

References

1 Breusch, R., On the sum of relative extrema of $|f(z)|$ on the unit circle, Bull. Amer. Math. Soc. 53 (1947), 982-986.

2 Clunie, J., On a problem of Erdös, J. London Math. Soc. 42 (1967), 133-136.

3 Danzer, L. W. and Ch. Pommerenke, Über die Diskriminante von Mengen gegebene Durchmessers, Monatsh. Math. 71 (1967), 100-113.

4 Elbert, A., Über eine Vermutung von Erdös betreffs Polynome I, II. Studia Sci. Math. Hungar. 1 (1966), 119-128 and 3 (1968), 299-324.

5 Erdös, P. An extremum problem concerning tingonometric polynomials, Acta Sci. Math. (Szeged) 9 (1939), 113-115.

6 Erdös, P., Some remarks on polynomials, Bull. Amer. Math. Soc. 53 (1947), 1169-1176.

7 =(I) Erdös, P., F. Herzog and G. Piranian, Metric properties of polynomials, J. d'Analyse Math. 6 (1958), 125-148.

8 Erdös, P. and E. Netanyahu, A remark on polynomials and the transfinite diameter, Israel J. Math. 14 (1973), 23-25.

9 Goodman, A. W. and R. E. Goodman, A circle covering theorem, Amer. Math. Monthly 52 (1945), 494-498.

10 Kristiansen, G. K., Proof of an inequality for trigonometric polynomials, Proc. Amer. Math. Soc., 44 (1974), 49-57.

11 Kristiansen, G. K., Proof of a polynomial conjecture, Proc. Amer. Math. Soc. 44 (1974), 58-60.

12 =(PI) Pommerenke, Ch., On some problems by Erdös, Herzog and Piranian, Michigan Math. J. 6 (1959), 221-225.

13 Pommerenke, Ch., On some metric properties of polynomials with real zeros, I and II, Michigan Math. J., 6 (1959), 377-380 and 8 (1961), 49-54.

14 =(PII) Pommerenke, Ch., On metric properties of complex polynomials, Michigan Math. J. 8 (1961), 97-115.

15 Saff, E. B. and T. Scheill-Small, Coefficient and integral mean estimates for algebraic and trignometric polynomials with restricted zeros, J. London Math. Soc. 9 (1974), 16-22.

P. Erdös
Hungarian Academy of Sciences, Budapest, Hungary

SOME NONLINEAR VARIATIONAL PROBLEMS

Stephen D. Fisher

We derive a necessary condition that a solution of the variational problem $\inf\{\|Tu\|: u \in U\}$ must satisfy and show how this condition can be used to solve certain nonlinear equations and to derive detailed properties of the solution of the minimum curvature problem.

1 Introduction and main results

Let X and Y be real Banach spaces, let U be a subset of X and let T be a continuous map of U into Y. We shall investigate the existence and nature of solutions of the variational problem

$$(1)\quad \alpha = \inf\{\|Tu\| : u \in U\}.$$

Of particular interest is the case when Y is $L^\infty(\Omega,\mu)$ or some closed subspace hereof and for this reason we shall in the following restrict ourselves to the situation when X and Y are dual spaces of separable Banach spaces W and Z, respectively. We shall also assume that we are given weak-* continuous Frechêt differentiable functionals $\ell_o,\ldots,\ell_m$ on X and real numbers $r_o,\ldots,r_m$ and that

$$(2)\quad U = \{x \in X: \ell_j(x) \le r_j,\ j = 0,\ldots,m\}.$$

If any ℓ_j is linear then the equality $\ell_j(x) = r_j$ may be used in the definition of U without any change in the conclusions. The following is the simplest existence theorem.

Preceding page blank

THEOREM 1. *If T is weak-*continuous and if there is a bounded minimizing sequence for* (1), *then there is a solution* Tu_o *of* (1).

In order to obtain information about the solution of (1) we assume that T is Frechét differentiable at u_o; that is, there is a bounded linear operator L from X into Y with

(3) $T(u_o + \delta x) = Tu_o + \delta Lx + 0(\delta^2)$, $\delta \to 0$

whenever $u_o + \delta x$ lies in the domain of T for all small enough δ.

THEOREM 2. *Let* $u_o \varepsilon U$ *be a solution of* (1), *let* J *consist of those indices for which* $\ell_j(u_o) = r_j$, *let* ℓ'_j *be the Frechét derivative of* ℓ_j at u_o *and set*

$$V = \{x \varepsilon X: \ell'_j(v) \leq 0 \text{ for all } j \varepsilon J\},$$

$$V_o = \{x \varepsilon X: \ell'_j(v) = 0 \text{ for all } j \varepsilon J\}.$$

Let L *be the Frechét derivative of* T *at* u_o *and suppose that* L *is weak-* continuous and that* LV_o *has finite codimension in* Y. *Let*

$$*V = \{m \varepsilon Z: m(Lv) \geq 0 \text{ for all } v \varepsilon V\}.$$

Then

(4) $\alpha = \inf\{\|Tu_o + Lv\| : v \varepsilon V\}$

and

(5) $\alpha = \max\{|m(Tu_o)| : \|m\| \leq 1, m \varepsilon *V\}$.

The proof of Theorem 2 is obtained by using (3) to establish (4) and then using duality to obtain (5) making crucial use of the fact that LV_o has finite codimension in Y to insure that the equality in (5) is actually attained for some $m \in {}^*V$. Note that if some ℓ_j is linear, then the equality $\ell_j(x) = 0$ may be used in the definition of V.

COROLLARY 3. *If* $Y = L^\infty(\Omega,\mu)$ *then there is an* $h \in L^1(\Omega,\mu)$ *with norm* 1 *and*

(i) $0 \leq \int_\Omega hLv,\ v \in V$

(ii) $0 \leq hTu_o$ a.e.μ

(iii) $|Tu_o| = \alpha$ a.e.μ where $h \neq 0$.

Proof. (i) follows from Theorem 2 while (ii) and (iii) are consequences of equality in Hölder's inequality.

COROLLARY 4. *The number* α *is the distance from* Tu_o *to* LV_o; *in particular, if there are no restraints, then* α *is the distance from* Tu_o *to* LX.

2 Applications to nonlinear equations

EXAMPLE 1. Let $F(t,\underline{x})$, $t \in [0,1]$, $\underline{x} \in \mathbb{R}^n$, be C^1 and consider the initial-value problem

$$(6)\quad y^{(n)}(t) = F(t,y(t),\ldots,y^{(n-1)}(t))$$

$$y^{(j)}(0) = r_j,\quad j = 0,\ldots,n-1.$$

Set $X = W^{n,\infty}(0,1)$, $Ty(t) = y^{(n)}(t) - F(t,y(t),\ldots,y^{(n-1)}(t))$, $\ell_j(y) = y^{(j)}(0)$ for $j = 0,\ldots,n-1$. Then $V = W_o^{n,\infty}(0,1)$ and the Frechét derivative of T at u_o is

$$Lg(t) = g^{(n)}(t) + \sum_{j=0}^{n-1} A_j(t)g^{(j)}(t)$$

where $A_o, \ldots, A_{n-1}$ are continuous functions depending on F and u_o. Hence, L maps V_o onto $Y = L^\infty(0,1)$. Therefore, in order to conclude that (6) has a solution we just need to impose conditions on $F(t,\underline{x})$ which assure that (1) has a solution. One such condition is that

$$|F(t,\underline{x})| \leq (1 - \delta)|\underline{x}| + M, \text{ some } \delta > 0, M > 0.$$

Another is

$$|F(t,\underline{x})| \leq C|x_n| + M, \text{ some } C,M > 0.$$

EXAMPLE 2. Let g be a smooth monotone increasing function on $\mathbb{R}$ with $|g(x)|/|x| \to 0$ as $x \to \pm\infty$. Let $f \in L^\infty(0,1)$ and consider the nonlinear boundary-value problem

(7) $$(-1)^K u^{(2K)}(t) + g(u(t)) = f(t)$$

$$u^{(\nu)}(0) = a_\nu, \; u^{(\nu)}(1) = b_\nu, \; \nu = 0,\ldots,K-1.$$

Here we put $X = W^{2K,\infty}(0,1)$, $Y = L^\infty(0,1)$, $\ell_\nu u = u^{(\nu)}(0)$ for $\nu = 0,\ldots,K-1$ and $\ell_\nu u = u^{(\nu-K)}(1)$ for $\nu = K,\ldots,2K-1$. Let

$$(Tu)(t) = (-1)^K u^{(2K)}(t) + g(u(t)) - f(t).$$

Then the Frechét derivative of T at u_o is

$$(Lv)(t) = (-1)^K v^{(2K)}(t) + g'(u_o(t))v(t)$$

which maps V_o onto L^∞ showing that (7) has a solution;

the growth condition on g guarantees that the minimization problem has a solution.

EXAMPLE 3. Let L_o be a given number and g a function in $L^p(0,L_o)$, $1 < p \leq \infty$. Then there is a curve, parametrized by arc-length, with length L_o and curvature g. That is, there is a pair of functions $x(t)$, $y(t) \in W^{2,p}(0,L_o)$ with

$$\text{(a)}\ \dot{x}^2(t) + \dot{y}^2(t) \equiv 1$$

$$\text{(b)}\ \dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t) = g(t), \quad 0 \leq t \leq L_o.$$

This is proved by letting U consist of all pairs (x,y) in $X = W^{2,p}(0,L_o) \oplus W^{2,p}(0,L_o)$ which satisfy (a) and setting $T(x,y) = g - \kappa(x,y)$ where $\kappa(x,y)$ is the curvature, given in this case by the left-hand side of (b).

3 Minimum curvature

As a last application of Theorem 2 and Corollary 3 we investigate the following problem: Let $\mathcal{P} = \{p_1,\ldots,p_N\}$ be a set of distinct points in the plane. We wish to pass a smooth curve $t \to (x(t),y(t))$ through the set $\mathcal{P}$ whose curvature, measured in the L^∞ norm, is as small as possible. We impose the constraint that the lengths of the competing curves have a uniform bound L_o for otherwise the infimum of the curvatures is zero and, except in the trivial case when all the points lie on a single straight line, there is no smooth curve with zero curavture passing through all the points. When the arc-length parametrization is used the curvature formula

$$T(x,y)(t) = (\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t))(\dot{x}(t)^2 + \dot{y}(t)^2)^{-3/2}$$

simplifies to

$$T^2 = \ddot{x}(t)^2 + \ddot{y}(t)^2$$

and hence every minimizing sequence is bounded in $X = W^{2,\infty}(0,L_o) \oplus W^{2,\infty}(0,L_o)$. Let (x,y) be a solution parametrized by arc-length and let points $0 \le t_1 \le \dots \le t_N \le L_o$ be chosen with $(x(t_j),y(t_j)) = p_j$, $j = 1,\dots,N$. We let $\ell_j(u,v) = (u(t_j),v(t_j))$ for $j = 1,\dots,N$ and

$$\ell_o(u,v) = \int_0^{L_o} (\dot{u}^2 + \dot{v}^2)^{1/2}$$

so that the nonlinear constraint is $\ell_o(u,v) \le L_o$. The Frechét derivative of T at (x,y) is given by

$$L(u,v) = -\dot{v}\ddot{x} + \ddot{v}\dot{x} + \dot{u}\ddot{y} - \ddot{u}\dot{y} - 3(\dot{x}\dot{u} + \dot{y}\dot{v})\kappa$$

where κ is the curvature of (x,y). It is not hard to see that L maps X onto L^∞; thus, according to Theorem 2 and Corollary 3 there is a function $h \in L^1$ with $0 \le \int hL(u,v)$ whenever $u(t_j) = v(t_j) = 0$ for $j = 1,\dots,N$ and $0 \ge \ell_o'(u,v)$; this h also satisfies $\kappa h \ge 0$ a.e., and $|\kappa| = \alpha$ a.e. where $h \ne 0$. Some integration by parts shows that h satisfies the differential equation

$$\text{(9)} \quad \ddot{h} + \kappa^2 h = \lambda\kappa$$

where λ is a scalar, $\lambda \le 0$; if $\lambda < 0$, then whenever $h = 0$, we must have $\kappa = 0$ also. In general, if h is, say, positive on (a,b) with $h(a) = h(b) = 0$ then (9) shows that $b - a \ge \pi/\alpha$. Thus the intervals on which h has one sign have, with at most 2 exceptions, length at least π/α. If $\lambda = 0$, then it is possible that h vanishes on some interval (a,b) and yet κ does not. If so we may just repeat the

minimization process on (a,b) using as admissible functions those pairs (ξ,η) in U which agree with (x,y) outside the interval (a,b). In this way we obtain a solution consisting of finitely many arcs of circles and straight line segments.

S. D. Fisher
Northwestern University and Mathematics Research Center, University of Wisconsin. Research supported in part by the National Science Foundation and by U.S. Army Contract No. DAAG29-75-C-0024

THE PROBLEM OF MINIMAL PROJECTIONS IN $\mathcal{L}_1$-SPACES

C. Franchetti and E. W. Cheney

This note presents an outline of the theory of minimal projections onto finite-dimensional subspaces of $\mathcal{L}_1(T,\Sigma,\mu)$. We define a <u>projection</u> to mean a bounded linear map $P: X \to Y$ of a normed space X onto a subspace Y having the property that $Py = y$ for all $y \in Y$. Such a projection is <u>minimal</u> if $\|P\| \leq \|Q\|$ for every projection Q from X onto Y. A concrete determination of the minimal projection of $\mathcal{L}_1[-1,1]$ onto the polynomials of degree 1 is given as an illustration of the theory.

If P is a projection of a normed space X onto a finite-dimensional subspace Y, then for any basis $\{y_1,\ldots,y_n\}$ of Y there exist functionals $f_1,\ldots,f_n$ in X^* such that $P = \sum_{i=1}^{n} f_i \otimes y_i$. This notation means $Px = \sum_{i=1}^{n} f_i(x)y_i$. If $X = \mathcal{L}_1(T,\Sigma,\mu)$ and if (T,Σ,μ) is a σ-finite measure space, then X^* can be identified with $\mathcal{L}_\infty(T,\Sigma,\mu)$ and the projection takes the form $Px = \sum_{i=1}^{n} (x,u_i)y_i$ where $u_i \in \mathcal{L}_\infty$ and $(x,u) = \int x(t)u(t)\,d\mu(t)$. The projection property is equivalent to the requirement that $(u_i,y_j) = \delta_{ij}$ $(1 \leq i,j \leq n)$.

For the projection $P = \sum u_i \otimes y_i$ we define the kernel to be $K(s,t) = \sum_{i=1}^{n} u_i(t)y_i(s)$. The projection has the form of an integral operator with kernel K: $(Px)(s) = \int K(s,t)x(t)dt$. Next we define the Lebesgue function of P to be

$$\Lambda_P(t) = \int |K(s,t)|\,ds.$$

One can prove that Λ depends only on P and not on its particular representation. Furthermore, the norm of P as an operator on $\mathcal{L}_1$ is given by

Preceding page blank

$\|P\| = \text{ess sup } \Lambda_P(t)$.

The minimal projection problem for the subspace having basis $\{y_1,\ldots,y_n\}$ can now be stated as follows. Among the n-tuples $(u_1,\ldots,u_n)$ satisfying $u_i \in \mathcal{L}_\infty$ and $(u_i,y_j) = \delta_{ij}$, find one for which the expression $\text{ess}_t\text{sup} \int |\sum u_i(t)y_i(s)|ds$ is a minimum.

The first important characteristic of a minimal projection in this setting is that its Lebesgue function must be constant. The precise theorem is as follows.

THEOREM I. Let (T,Σ,μ) be a finite and nonatomic measure space. Let P be a minimal projection of $\mathcal{L}_1(T,\Sigma,\mu)$ onto a smooth finite-dimensional subspace. Then Λ_P is constant.

Here the word "smooth" is used as in Banach space theory: each point on the unit sphere has a unique supporting hyperplane. This is equivalent to the requirement that no non-zero function in the subspace can vanish on a set of positive measure.

Another necessary condition satisfied by a minimal projection is as follows.

THEOREM II. Assume the same hypotheses as in Theorem I, and let $P = \sum_{i=1}^{n} u_i \otimes y_i$ be a minimal projection. Then there do not exist $v_1,\ldots,v_n \in Y^\perp \cap \mathcal{L}_\infty$ satisfying

$$\text{ess}\sup_t \sum_{i=1}^{n} v_i(t)f_i(t) < 0.$$

where $f_i(t) = \int y_i(s)\text{sgn } K(s,t)ds$.

THEOREM III. The two necessary conditions given in Theorems I and II taken together form a sufficient condition for the minimality of a projection.

In order to give an idea of the techniques used in proving these theorems, we shall outline here the proof of Theorem I.

Suppose that P is a projection whose Lebesgue function Λ is not constant. Then there exists a positive number δ such that the two sets

$$A = \{t : \Lambda(t) \geq \|P\| - \delta\}$$

$$B = \{t : \Lambda(t) < \|P\| - \delta\}$$

have positive measure. In a separate lemma we establish that there exist disjoint measurable sets $B_1, \ldots, B_n$ in B and a basis $\{y_1, \ldots, y_n\}$ for Y such that $\int_{B_i} y_j = \delta_{ij}$. This requires that the measure space be nonatomic. Next, select $u_1, \ldots, u_n$ in $\mathcal{L}_\infty$ such that $P = \sum u_i \otimes y_i$. We need now the fact that Λ is independent of the representation of P, so that $\Lambda(t) = \|\sum u_i(t) y_i\|_1$. Now we perturb the functions u_i as follows.

$$\tilde{u}_i(t) = \begin{cases} u_i(t) + \epsilon_{ij} & \text{if } t \in B_j \\ (1-\epsilon) u_i(t) & \text{if } t \in C. \end{cases}$$

Here $C = T \setminus (B_1 \cup \ldots \cup B_n)$ and $\epsilon_{ij} = \epsilon \int_C u_i y_j$. Because of the special form of the perturbed u_i, one can show that the operator $Q_\epsilon = \sum \tilde{u}_i \otimes y_i$ is a projection onto Y and that $\|Q_\epsilon\| < \|P\|$ for some ϵ.

As an example of the theory, we now outline a determination of the minimal projection from $\mathcal{L}_1[-1,1]$ onto π_1 the space of 1st-degree polynomials. This problem was posed to us by Professor Carl de Boor. One begins by calculating the $\mathcal{L}_1$-norm of an arbitrary element of π_1. Namely, $\int_{-1}^{1} |a+bt|dt = 2|a|$ or $(a^2+b^2)/|b|$ according to whether $|a| \geq |b|$ or $|a| < |b|$. Next we fix a basis for the subspace: $y_1(t) = 1$ and $y_2(t) = t$. It is possible then to prove that among the minimal projections $P = \sum_{i=1}^{2} u_i \otimes y_i$ there is one for which u_1 is even and u_2 is odd. We next show that if such a projection has the additional properties

$$0 < |u_1| \le |u_2|$$

$$u_2^2 - u_1^2 = 2\lambda t u_1 u_2$$

$$u_1^2 + u_2^2 = \rho |u_2|$$

for appropriate constants λ and ρ, then it must be minimal. The two functions which have all these properties are given on the interval $[0,1]$ by the formulas

$$u_2(t) = \frac{-\lambda}{2 \log \xi \left(1 + \lambda^2 t^2 - \lambda t \sqrt{\lambda^2 t^2 + 1}\right)}$$

$$u_1(t) = u_2(t)\left[-\lambda t + \sqrt{\lambda^2 t^2 + 1}\right]$$

in which $\lambda = (1 - \xi^2)/(2\xi)$ and ξ is the root of $2\xi(1 - \xi^2 + \xi)\log \xi + 1 = \xi^2$ between 0 and 1. The numerical values are $\xi = .32796779$, $\lambda = 1.36055608$, $\|P\| = -\lambda/\log \xi = 1.22040491$. The minimal projection of $\mathcal{L}_1$ onto π_1 is unique.

C. Franchetti*
Department of Mathematics
The University of Texas
at Austin,
Austin, Texas 78712

E. W. Cheney*
Department of Mathematics
The University of Texas
at Austin
Austin, Texas

*Dr. Franchetti was supported by a NATO Science Fellowship administered through the National Research Council (CNR) of Italy; Dr. Cheney was supported in part by the National Science Foundation, Grant No. MPS75-07939.

MARKOV-BERNSTEIN TYPE INEQUALITIES IN $L_p(-\infty,\infty)$

Geza Freud

Let $w(x) = \exp\{-Q(x)\}$ where $Q(x)$ is an even convex function defined on $(-\infty,\infty)$. We define the sequence $\{q_n\}$ by the relation

$$q_n Q'(q_n - 0) \leq n \leq q_n Q'(q_n + 0) \quad (n = 1,2,\ldots)$$

and we set $q_0 = q_1$.

Then under suitable additional conditions on Q we have

THEOREM 1. *For an arbitrary polynomial* τ_n *of degree not greater than* n *we have for every* $1 \leq n \leq \infty$

$$(*) \quad \|\tau_n\|_p \leq A \frac{n}{q_n} \|\tau_n'\|_p \ .$$

Here $\|\cdot\|_p$ *is the norm of* $L_p(-\infty,\infty)$. *The number* A *depends on the choice of* w *but does not depend on* n *or* τ_n.

In particular, the following conditions (i), (ii), (iii) *together are sufficient for* (*) *to hold:*

(i) $Q \in C_2(0,\infty)$,

(ii) $Q''(x)$ is positive and nondecreasing in $(0,\infty)$ and

(iii) $x \dfrac{Q''(x)}{Q'(x)} \leq c_0 \quad (0 < x < \infty)$.

Note that (i), (ii) and (iii) are satisfied if $\alpha \geq 2$ and $Q(x) = |x|^\alpha$, i.e., $w(x) = \exp\{-|x|^\alpha\}$.

We apply (*) to prove precise Bernstein-type converse theorems in weighted polynomial approximation.

For more special weights and using simpler arguments we prove similar results in [2] and [3].

1 Proof of the inequality (*)

The case $p = \infty$ of (*) will be proved separately in our paper [6]. The only comment we make here is that far deeper arguments are needed than those which applied to prove some special cases of it in our papers [1] and [4]. In the course of extending (*) to $1 \leq p \leq \infty$ we will need the following reseults proved elsewhere: We denote by $\{s_\nu(w^2;f;x)\}$ the partial sums of the expansion of a function f satisfying $fw \in L_p$ in the orthonormal polynomial system $\{p_\nu(w^2;x)\}$ and we introduce the delated de la Vallée Poussin means

(1.1) $$v_n(w^2;f;x) = \frac{1}{n}\sum_{n+1}^{2n} s_\nu(w^2;f;x).$$

Then by [6] Lemma 4.1 resp. [3] Theorem 4.1 we have

(1.2) $$\|v_n(w^2;f;\cdot)w\|_\infty \leq c_1\|fw\|_\infty$$

and in virtue of [6] Theorem 3.1

(1.3) $$\|v_n'(w^2;f;.)w\|_\infty \leq c_2 \frac{n}{q_n} \|fw\|_\infty .$$

Let us denote by P_n the set of polynomials which have degrees not greater than n. Then we have

(1.4) $$v_n(w^2;\tau_n) = \tau_n, (\tau_n \in P_n)$$

and for arbitrary f and every $\tau_n \in P_n$ we have

(1.5) $$\int[f - v_n(w^2;f)]\tau_n w^2 \, dx = 0.$$

In order to prove (*) for $p = 1$ we observe that

$$(1.6) \quad \|\tau_n' w\|_1 = \sup \int \tau_n' \, g \, w^2 \, dx$$

where g runs through all functions satisfying $\|gw\|_\infty \leq 1$.

By (1.5) and a subsequent integration by parts

$$(1.7) \quad \int \tau_n' g w^2 dx = \int \tau_n' v_n(w^2;g) w^2 dx = -\int \tau_n v_n'(w^2;g) w^2 dx$$

$$+ 2 \int \tau_n Q' v_n(w^2;g) w^2 dx$$

so that

$$(1.8) \quad \left|\int \tau_n' g w^2 dx\right| \leq \|\tau_n w\|_1 [\|v_n'(w^2;g;\cdot)w\|_\infty + 2\|Q' v_n(w^2;g;\cdot)w\|_\infty].$$

The first term in brackets can be estimated by (1.3). As to the second term, we have in virtue of [5] Lemma 4.2 and (1.2), (1.3),

$$(1.9) \quad \|Q' v_n(w^2;g;\cdot)w\|_\infty \leq \|Q'[v_n(w^2;g;.) - v_n(w^2;g;0)]w\|_\infty$$

$$+ c_3 |v_n(w^2;g;0)w(0)| \leq c_4 \|v_n'(w;g;\cdot)w\|_\infty$$

$$+ c_3 \|v_n(w^2;g;.)w\|_\infty \leq c_5 \frac{n}{q_n} \|gw\|_\infty \leq c_5 \frac{n}{q_n} \quad .$$

We infer from (1.6) that

$$\|\tau_n' w\|_1 \leq (c_2 + c_5) \frac{n}{q_n} \|\tau_n w\|.$$

We know that (*) is valid for $p = 1$ and $p = \infty$. The general case $1 \leq p \leq \infty$ can be obtained from these two particular cases by applying the Riesz-Thorin interpolation theorem. (Same argument as in [1] and [4]).

2 Remarks on direct theorems of weighted polynomial approximation

In our paper [3] we introduce generalized L_p-continuity moduli of functions with respect to weights as follows. For a Φ satisfying $\Phi w \in L_p$ we set

$$(2.1)\quad \omega(L_p, w; \Phi, \delta) = \sup_{|h| \le \delta} \|T_h(\Phi w) - \Phi w\|_p + \delta \|Q'_\delta \Phi w\|_p$$

where T_h is the translation operator, i.e., $T_h[f(x)] = f(x+h)$ and

$$(2.2)\quad Q'_\delta(x) = \min[\delta^{-1}, |Q'(x)|].$$

Note that

$$(2.3)\quad 0 \le \delta Q'_\delta(x) \le 1 \quad (\delta \ge 0).$$

This expression characterizes the order of optimal weighted polynomial approximation in the same way as the usual continuity modulus of 2π-periodic functions determines the order of trigonometric approximation. In fact let

$$(2.4)\quad \varepsilon_n^{(p)}(w; f) = \inf_{\tau \in P_n} \|(f - \tau)w\|_p \;;$$

then, as we proved in [3]

$$(2.5)\quad \varepsilon_{n+r}^{(p)}(w; \Phi) \le c_6 e^{c_7 r} \left(\frac{q_n}{n}\right)^r \omega\left(L_p, w; \Phi^{(r)}, \frac{q_n}{n}\right).$$

This is the analogue of Jackson's trigonometric approximation theorem. In the last two chapters of [3] we deduced (2.5) from the two assumptions that (1.2) is valid and that

$$(2.6)\quad \varepsilon_n^{(1)}(w;\Phi) \leq c_8 \frac{q_n}{n} \int_{-\infty}^{\infty} w(x)\,|d\Phi(x)|.$$

Let us now observe that (1.2) certainly holds if (i), (ii) and (iii) are satisfied and that we proved in [5] that (2.6) is true if (i) is valid and we have

$$(2.7)\quad 1 < \varliminf_{x\to\infty} \frac{Q'(2x)}{Q'(x)} < \varlimsup_{x\to\infty} \frac{Q'(2x)}{Q'(x)} < \infty.$$

In what follows we replace (2.7) by the slightly more restrictive condition

$$(2.8)\quad 1 < c_9 < x\,\frac{Q''(x)}{Q'(x)} < c_{10}.$$

Thus we assume that (i), (ii) and (2.8) are satisfied. This is certainly true if $w(x) = \exp\{-|x|^{\alpha}\}$, $\alpha \geq 2$.

We observed in [5] that we can replace $\omega(L_p,w;\Phi^{(r)}, \frac{q_n}{n})$ by the smaller factor $\Omega(L_p,w;\Phi, \frac{q_n}{n})$ where

$$(2.9)\quad \Omega(L_p,w;f,\delta) = \inf_{a\varepsilon R} (L_p,w;f-a,\delta)$$

i.e., $\Omega(L_p,\omega,f,\delta) = \sup_{|h|\leq\delta} \|T_h(fw) - fw\|_p + \delta \inf_{a\varepsilon R}\|Q'_{\delta}(f-a)w\|.$

In fact, $\varepsilon_{n+r}^{(p)}(w;\Phi)$ does not change if we replace Φ by $\Phi - (1/r.)ax^r$. Consequently we can replace on right side of (2.5) $\omega(\Phi^{(r)})$ by $\inf\omega(\Phi^{(r)}-a)$. We can give a more accurate picture on weighted polynomial approximation in terms of Ω than in terms of ω. In particular, we introduced in [5] the Peetre type K functional

$$(2.10)\quad K(L_p,w;\Phi,\delta) = \inf_{f_1+f_2=\Phi} \{\|f_1w\|_p + \delta\|f_2'w\|_p\}$$

and proved that

$$(2.11)\quad c_{11}K(L_p,2;\Phi,\delta) \leq \Omega(L_p,w;\Phi,\delta) \leq c_{12}K(L_p,w;\Phi,\delta).$$

As a last remark to this chapter, let us observe that we have

$$(2.12)\quad \Omega(L_p,w;f_1+f_2,\delta) \leq \Omega(L_p,w;f_1,\delta) + \Omega(L_p,w;f_2,\delta).$$

3 The converse theorems of weighted polynomial approximation

On combining (*) with (2.11) and (2.10) we get for arbitrary $\tau_n \in P_n$

$$(3.1)\quad \Omega(L_p,w;\tau_n,\delta) \leq c_{12}K(L_p,w;\tau_n,\delta) \leq c_{12}\delta\|\tau_n'w\|_p$$
$$\leq Ac_{12}\delta\,\frac{n}{q_n}\,\|\tau_n w\|_p \quad (\tau_n \in P_n).$$

Let $\Phi w \in L_p$ and $\{\tau_k \in P_k\}$ be a polynomial sequence for which

$$(3.2)\quad \|(\Phi - \tau_k)w\|_p \leq 2\varepsilon_k^{(p)}(w;\Phi).$$

In view of (2.12) we have then

$$(3.3)\quad \Omega(L_p,w;\tau_{2^n-1},\delta) \leq \sum_{k=1}^{n} \Omega(L_p,w;\tau_{2^k-1} - \tau_{2^{k-1}-1};\delta)$$
$$\leq Ac_{12}\delta \sum_{k=1}^{n} \frac{2^k}{q_{2^k}} \|(\tau_{2^k-1} - \tau_{2^{k-1}-1})w\|$$
$$\leq Ac_{12}\delta \sum_{k=1}^{n} \frac{2^k}{q_{2^k}} [\|(\Phi - \tau_{2^k-1})w\|_p$$
$$+ \|(\Phi - \tau_{2^{k-1}-1})w\|_p]$$
$$\leq 2Ac_{12}\delta \sum_{k=1}^{n} \frac{2^k}{q_{2^k}} [\varepsilon_{2^k-1}^{(p)}(w;\Phi)$$

$$+ \varepsilon^{(p)}_{2^{k-1}-1}(w;\Phi)]$$

$$\leq c_{13}\delta \sum_{k=0}^{n-1} \frac{2^k}{q_{2^k}} \varepsilon^{(p)}_{2^k-1}(w;\Phi).$$

It follows by (2.10), (2.11) and (2.12)

$$(3.4) \quad \Omega(L_p,w;\Phi,\delta) \leq \Omega(L_p,w;\Phi - \tau_{2^n-1},\delta) + \Omega(L_p,w;\tau_{2^n-1},\delta)$$

$$\leq c_{12}\|(\Phi - \tau_{2^n-1})w\|_p + \Omega(L_p,w;\tau_{2^n-1}\delta)$$

$$\leq 2c_{12}\,\varepsilon^{(p)}_{2^n-1}(w,\Phi) + c_{13}\delta \sum_{k=0}^{n-1} \frac{2^k}{q_{2^k}}\varepsilon^{(p)}_{2^k-1}(w;\Phi).$$

THEOREM 3.1 We have

$$(3.5) \quad \Omega(L_p,w;\Phi,\frac{q_{2^n}}{2^n}) \leq c_{14}\frac{q^{2^n}}{2^n} \sum_{k=0}^{n-1} \varepsilon^{(p)}_{2^k-1}(w;\Phi).$$

Proof. Combining the identity

$$\frac{2q_n}{q_{2n}} = \exp\{\int_{q_n}^{q_{2n}} \frac{Q''(x)}{Q'(x)}\,dx\}$$

with our assumption (2.8) we obtain that

$$(3.6) \quad 1 < c_{15} < \frac{q_{2n}}{q_n} < c_{16} < 2\ ;$$

consequently

$$(3.7) \quad 2 > \frac{2}{c_{15}} > \frac{2^{k+1}}{q_{2^{a+1}}} / \frac{2^k}{q_{2^k}} > \frac{2}{c_{16}} > 1.$$

It follows

$$(3.8) \quad \varepsilon^{(p)}_{2^n-1}(w;\Phi) \leq c_{17}\frac{q_{2^n}}{2^n} \sum_{k=0}^{n-1} \frac{2^k}{q_{2^k}}\varepsilon^{(p)}_{2^n-1}(w;\Phi)$$

$$\leq c_{17} \frac{q_{2^n}}{2^n} \sum_{k=0}^{n-1} \frac{2^k}{q_{2^k}} \varepsilon^{(p)}_{2^k-1}(w;\Phi).$$

Inserting $\delta = 2^{-n} q_{2^n}$ in (3.4) and applying (3.8) we get (3.5), q.e.d.

THEOREM 3.2 <u>Let</u> $\Phi w \varepsilon L_p$ <u>and for a natural</u> r <u>let</u>

$$(3.9) \qquad \sum_{\nu=1}^{\infty} \nu^{r-1} q_\nu^{-r} \varepsilon_\nu^{(p)}(w;\Phi) < \infty.$$

<u>Then</u> Φ <u>has an</u> r-th <u>derivative</u> $\Phi^{(r)}$ <u>satisfying</u> $\Phi^{(r)} w \varepsilon L_p$ <u>and</u>

$$(3.10) \qquad \varepsilon^{(p)}_{n-r}(w,\Phi^{(r)}) \leq c_{18} e^{c_{19} r} \sum_{m=0}^{\infty} \left(\frac{2^m n}{q_{2^m n}}\right)^r \varepsilon^{(p)}_{2^m n}(w;\Phi).$$

<u>Proof</u>. Observe that (3.9) is equivalent to the convergence of the series on the right of (3.10). By (3.2) and (*) we have, considering also (3.7),

$$\|(\tau^{(r)}_{2^{m+1}n} - \tau^{(r)}_{2^m n})w\|_p \leq c_{18} e^{c_{19} r} \left(\frac{2^m n}{q_{2^m n}}\right)^r \varepsilon^{(p)}_{2^m n}(w;\Phi).$$

Summation over m gives the estimate (3.10) for $\|(\Phi^{(r)} - \tau_n^{(r)})w\|_p$, q.e.d.

THEOREM 3.3 <u>If for some nonnegative interger</u> r <u>and some</u> $0 < \rho < 1$ <u>there exists a constant</u> $B > 0$ <u>for which</u>

$$(3.11) \qquad \varepsilon_n^{(p)}(w;\Phi) \leq B\left(\frac{q_n}{n}\right)^{r+\rho}$$

<u>then</u> Φ <u>has an</u> r-th <u>derivative satisfying</u> $\Phi^{(r)} w \varepsilon L_p$ <u>and</u>

$$(3.12) \qquad \Omega(L_p, w; \Phi^{(r)}, \delta) \leq C(w,r,\rho) B \delta^\rho.$$

Note that if we replace in (3.12) Ω by ω then the factor C would depend on Φ.

References

1 Freud, G., On direct converse theorems of weighted polynomial approximation. Mathem. Zeitschrift 126 (1972), 123-124.

2 Freud, G., Investigations on weighted approximation by polynomials, Studia Sci. Math. Hung., 8 (1973), 285-305.

3 Freud, G., On polynomial approximation with respect to general weights, Springer Lecture Notes in Mathematics, 399 (1974), 149-179.

4 Freud, G., On converse theorems of weighted polynomial approximation, Acta Math. Ac. Sci. Hung., 24 (1974), 389-397.

5 Freud, G., Weighted polynomial approximation and K-functionals, in Theory of Approximation with Applications, A. G. Law and B. N. Sahney, eds., Academic Press, New York, 1976, 9-23.

6 Freud, G., On Markov-Bernstein type inequalities and their applications, Journal of Approximation Theory, to be published.

G. Freud
Department of Mathematics
The Ohio State University
Columbus, Ohio 43210

A SEMI-INFINITE LINEAR PROGRAMMING PROCEDURE AND APPLICATIONS TO APPROXIMATION PROBLEMS IN OPTIMAL CONTROL

K.-H. Hoffmann and A. Klostermair

We generalize the well known Dual Simplex Method to a convergent exchange algorithm for semiinfinite linear programs in $\mathbb{R}^n$. Neither for practicability nor convergence proof do we need some kind of Haar's Condition. The procedure differs from the Carasso-Laurent-Töpfer [1] and the Schäfer-Algorithm [4] in milder assumptions for convergence and, because we do not need to solve auxiliary problems at single iteration steps it seems to converge faster. Often approximation problems can be reformulated as semiinfinite programs in $\mathbb{R}^n$. Therefore the procedure seems appropriate for solving approximation problems under side conditions.

1 Semiinfinite convex programs

Let T_o, T_1 be disjoint index sets and $T := T_o \cup T_1$
For $t \in T$ we consider fixed elements

$$b_t \in \mathbb{R}, \quad y_t \in \mathbb{R}^n,$$

and define the hyperplane

$$H_t := \{x \in \mathbb{R}^n \mid \langle y_t, x \rangle = b_t\}$$

and the halfspace

$$HS_t := \{x \in \mathbb{R}^n \mid \langle y_t, x \rangle \geq b_t\},$$

where $\langle \cdot,\cdot \rangle$ means the innner product in $\mathbb{R}^n$. The set of feasible points for the programming problem is defined by

$$M := \left(\bigcap_{t \in T_o} H_t\right) \cap \left(\bigcap_{t \in T_1} HS_t\right).$$

Let $u \in \mathbb{R}^n$ be a fixed target vector. We consider the following semiinfinite linear programming problem

Preceding page blank

(P) Find $\underline{x} \in M$ and $\underline{a} \in \mathbf{R}$ such that

$$< u,\underline{x} > \; = \; \inf \{< u,x > \mid x \in M\} \; =: \underline{a} \; .$$

This is a linear program with a finite number of variables but an infinite number of side conditions. Approximation problems can be reformulated in problems of this type; e. g.: for a given function $f \in C\,[a,b]$ find an element

$$\tilde{v} \in V := \text{span}\,(v_1, \ldots, v_{n-1}) \subset C\,[a,b] \;\; \text{such that}$$

$$\bigwedge v \in V \;\; \|f - \tilde{v}\|_\infty \;\leq\; \|f - v\|_\infty$$

is equivalent to the semiinfinite linear program: find cofficients $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbf{R}^n$ such that α_n is minimal and

$$\pm\,(f(t) - \sum_{i=1}^{n-1} \alpha_i\, v_i(t)) \leq \alpha_n$$

for all $t \in [a,b]$.

Corresponding to (P) we define its dual program. Let

$$W := \{w = (w_t)_{t \in T} \mid w_t \in \mathbf{R},\; w_t \neq 0 \;\; \text{for } t \in T' \subset T, |T'| < \infty\},$$

The set of dual feasible points is defined by

$$M^* := \{w \in W \mid w_t \geq 0 \;\; \text{for} \;\; t \in T_1, \;\; \sum_{t \in T} w_t y_t = u\} \; .$$

(D) Find $\bar{w} \in M^*$ and $\bar{a} \in \mathbf{R}$ such that

$$\sum_{t \in T} \bar{w}_t b_t = \sup \{ \sum_{t \in T} w_t b_t \mid w \in M^*\} =: \bar{a}.$$

This linear programm has a finite number of side conditions but an infinite number of variables.

For a geometric interpretation we introduce the notation

$$R_i := \{(y_t, b_t) \in \mathbf{R}^{n+1} \mid t \in T_i \}, \quad i=0,1 \quad \text{(Restriction set)}$$

and $R := R_1 \cup R_2$. Then x is a feasible point for (P) iff

$$R_o \subset H(x,-1) := \{(y,b) \in \mathbf{R}^n \times \mathbf{R} \mid < y,x > - b = 0\} \quad \text{and}$$

$$R_1 \subset HS(x,-1) := \{(y,b) \in \mathbf{R}^n \times \mathbf{R} \mid < y,x > - b \geq 0\} \; .$$

Let $G := \{u\} \times \mathbf{R} \subset \mathbf{R}^{n+1}$. Then (P) is equivalent to

(PG): Find a hyperplane $\underline{H}(x,-1)$ such that

$$R_o \subset \underline{H}(x,1), \quad R_1 \subset \underline{HS}(x,-1) \quad \text{and}$$

$$H(x,-1) \cap G = \{(u^{(1)},\ldots,u^{(n)}, \underline{a})\} \quad \text{with } \underline{a} \text{ minimal.}$$

Define

$$C := \{ \sum_{t \in T} w_t \binom{y_t}{b_t} \mid w_t \in \mathbf{R} \text{ for } t \in T, \quad w_t \geq 0 \text{ for } t \in T_1 \},$$

then (D) is equivalent to

(DG): Find a point $\bar{P} := (y^{(1)},\ldots,y^{(n)},\bar{a}) \in G \cap C$ with $\bar{a}$ maximal.

2 Relations between problem (P) and (D)

The exchange algorithm presented in this paper solves problem (D). We state relations between (P) and (D) in the following strong duality theorem.

THEOREM 1. _Let_ u _satisfy the condition_

(V) u _is relative interior point of_ C_o,

$$C_o := \{y \in \mathbf{R}^n \mid \bigvee b \in \mathbf{R} \ (y,b) \in C\}.$$

Then: $\underline{a} = \bar{a}$ _(strong duality property) and in the case_ $M \neq \emptyset$ _a minimal point_ $\underline{x}$ _of_ (P) _exists._

The main idea in the proof is a separation argument for suitable defined convex sets in $\mathbf{R}^{n+1}$. Condition (V) seems to be very strong but in the case $M \neq \emptyset$ and span $\{y_t \mid t \in T\} = \mathbf{R}^n$ it is equivalent to $\underline{M} := \{\underline{x} \in M \mid \langle u,x \rangle = \underline{a}\} \neq \emptyset$ and bounded.

For the construction of the algorithm and as a stopping condition on the iteration we need the following characterization theorem.

THEOREM 2. _Let_ (P) _and_ (D) _have the strong duality property and let_ $\underline{x}$ _be a feasible point for_ (P).

Let $T^* := \{t_1, t_2, \ldots, t_k\}$ be a finite subset of T and $w_1^*, w_2^*, \ldots, w_k^* \in \mathbb{R}$ with $w_i^* > 0$ for $t_i \in T_1$. We define:

$$y_i^* := y_{t_i}, \quad b_i^* := b_{t_i} \quad \text{and}$$

$$\bar{w}_t := \begin{cases} w_i^* & \text{for } t = t_i \quad (1 \leq i \leq k), \\ 0 & \text{for } t \notin T^*. \end{cases}$$

Let $\sum_{i=1}^{k} w_i^* y_i^* = u.$

Then the following statements are equivalent:

(i) $\underline{x}$ solves (P) and $\bar{w} = (\bar{w}_t)_{t \in T}$ solves (D).

(ii) $\langle u, \underline{x} \rangle = \sum_{t \in T} \bar{w}_t b_t$

(iii) $\langle y_i^*, \underline{x} \rangle = b_i^*$ for $i = 1, 2, \ldots, k.$

We intend to construct a maximal sequence of feasible points for (D). Similar to finite programs it is sufficient to consider a subset of M^* consisting of basic solutions.

DEFINITION 3. Let $w = (w_t)_{t \in T} \in W$ and $T(w) := \{t \in T \mid w(t) \neq 0\}$.

(i) $w \in W$ is a basic solution $\Leftrightarrow$ $(y_t)_{t \in T(w)} \subset \mathbb{R}^n$ linearly independent.

(ii) $w \in W$ is a feasible basic solution $\Leftrightarrow$ $w \in M^*$, w is a basic solution.

THEOREM 4. If (D) has a solution, there exists a basic solution solving (D).

DEFINITION 5. $w = (w_t)_{t \in T} \in W$ is degenerate $\Leftrightarrow$ $\operatorname{span}\{y_t \mid t \in T(w) \cup T_0\} \neq \mathbb{R}^n$.

3 The algorithm

In the step s we have a basic solution $\bar{w}^s$ and according to Definition 5 an index set $T(\bar{w}^s) \subset T$ with $(y_t)_{t \in T(\bar{w}^s)}$

linearly independent. In contrast to the Carasso-Laurent-Töpfer and the Schäfer Algorithm, at each step we work with an index $T^s \supset T(\bar{w}^s)$ with $(y_t)_{t \in T^s}$ linearly independent and $|T^s| = n$. For this purpose we introduce an auxiliary index set

$T_2 := \{1,2,\ldots,n\}$, $T \cap T_2 = \emptyset$, and vectors $y_t := e_t$ (t-th unit vector in $\mathbb{R}^n$), $b_t = 0$ for $t \in T_2$. If necessary we complete $(y_t)_{t \in T(\bar{w}^s)}$ to a basis of $\mathbb{R}^n$ by adding vectors y_t with $t \in T_2$. The property $|T^s| = n$ has the advantage that the linear equations we have to solve at each step have a unique solution, and therefore it is not necessary to solve auxiliary problems. On the other hand the maximal values a^s are not strong monotone increasing and it may be that cycles appear. To avoid this disadvantage we introduce auxiliary target vectors $(u_i^s)_{1 \le i \le n}$, linearly independent, such that the sequence $(a^s, a_1^s, \ldots, a_n^s)_{s \in \mathbb{N}}$ is strong monotone increasing in the sense of the lexicographical order.

In the s-th iteration we have the following basic situation:

for each $i \in I := \{1,2,..,n\}$ we have parameters $t_i^s \in T \cup T_2$ and for each $p \in I$ vectors $u_p^{(s)} \in \mathbb{R}^n$. For short we set

$I_k^s := \{i \in I \mid t_i^s \in T_k\}$, $k=0,1,2$;

$y_i^s := y_{t^s}$, $b_i^s := b_{t_i^s}$, $i \in I$;

$T^s := \{t_i^s \mid i \in I\}$.

(A1) $(y_i^s)_{i \in I}$ is a basis in $\mathbb{R}^n$;

(A1') $(u_p^s)_{p \in I}$ is a basis in $\mathbb{R}^n$.

For $i \in I$ exists $w_i^s \in \mathbb{R}$ with

$$\text{(A2)} \quad \begin{cases} w_i^s \ge 0 & \text{for } i \in I_1^s, \\ w_i^s = 0 & \text{for } i \in I_2^s. \end{cases}$$

(A3) $u = \sum_{i \in I} w_i^s y_i^s$.

AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1

APPROXIMATION THEORY II

NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.

UNCLASSIFIED AFOSR TR-76-1242 N/L

5 OF 7

AD
A033032

OF 7

3032

For $i \in I$, $p \in I$ there exists $w^s_{i,p} \in \mathbb{R}$ with

(A2)[1] $\bigwedge i \in I_1^s \bigvee 0 \leq p \leq n \; (w_{i,p} > 0 \wedge \bigwedge 0 \leq p' < p \quad w^s_{i,p'} = 0)$;

(A3) $\bigwedge p \in I \quad u^s_p = \sum_{i \in I} w^s_{i,p} \, y^s_i$.

Now we define index sets

$$J^s_p := \{ i \in I^s_1 \mid w^s_{i,p} > 0 \wedge \bigwedge 0 \leq p' < P \quad w^s_{i,p'} = 0 \} .$$

$\{J^s_p\}_{0 \leq p \leq n}$ is a disjoint partition of I^s_i.

(A4) $\bar{w}^s_t := \begin{cases} w^s_i & \text{for } t = t_i, \; i \in I^s_0 \cup I^s_1 , \\ 0 & \text{else} \end{cases}$,

(A5) $\bar{w}^s := (\bar{w}^s_t)_{t \in T}$ is a feasible point for (D).

Starting from this basic situation we proceed in the following way:

according to the characterization theorem 2, we solve the linear equations

(A6) $\bigwedge i \in I \quad < y^s_i, x^s > \; = b^s_i$

and set $a^s := \Sigma \, \bar{w}^s_t \, b_t = < u, x^s >$.

Now compute the quantity

$$g(x^s) := \sup(\{ |b_t - < y^t, x^s >| \mid t \in T_0 \} \cup \{ b_t - < y_t, x^s > \mid t \in T_1 \})$$

and distinguish

1. $g(x^s) \leq 0$, then x^s is a solution of (P) and $\bar{w}^s$ is a solution of (D) according to theorem 2,
2. $g(x^s) > 0$.

Then x^s is no feasible point. We select an exchange parameter $t^{*s} \in T$ such that

$$d^s := v^s(b_{t^{*s}} - < y_{t^{*s}}, x^s >) > 0,$$

---

1) $w^s_{i,o} := w^s_i$

where

$$v^s := \begin{cases} \operatorname{sign}(b_{t^{*s}} - \langle y_{t^{*s}}, x^s \rangle) & \text{for } t^{*s} \in T_0, \\ 1 & \text{for } t^{*s} \in T_1. \end{cases}$$

Because of condition (A1) the equations

$$\sum_{i \in I} c_i^s y_i^s = v\, y^{*s} \qquad , i \in I,$$

have a unique solution vector $(c_i^s)_{i \in I}$.

We distinguish two subcases:

a. $\bigwedge i \in I_1^s \quad c_i^s \leq 0 \wedge \bigwedge i \in I_2^s \quad c_i^s = 0.$

Then $\underline{a} = \bar{a} = \infty$ and (P) has no feasible points.

b. $\bigvee i \in I_1^s \quad c_i^s \quad 0 \vee \bigvee i \in I_2^s \quad c_i^s \neq 0.$

Then we replace a certain parameter $t^s \in I^s$ by a parameter $t^{*s} \in T \cup T_2$. For that define

$$r^{s+1} := \min \{\frac{w_i^s}{c_i^s} \mid (i \in I_1^s \wedge c_i^s > 0) \vee (i \in I_2^s \wedge c_i^s \neq 0)\}.$$

Now select an index

$$i^* := i^{*s} \in I'^s := \{i \in I_i^s \cup I_2^s \mid c_i^s \neq 0 \wedge \frac{w_i^s}{c_i^s} = r^{s+1}\}$$

and vector system $(u_p^{s+1})_{p \in I}$ such that conditions (A1) to (A3') are satisfied with s replaced by $s+1$ and

(1) $$t_i^{s+1} := \begin{cases} t_i^s & \text{for } i \in I \setminus \{i^{*s}\}, \\ t^{*s} & \text{for } i = i^{*s} \quad , \end{cases}$$

(2) $\bigwedge 1 \leq p' \leq p^{s} \quad u_{p'}^{s+1} := u_{p'}^s$, where

$$p^{*s} := \begin{cases} p & \text{for } i^{*s} \in J_p^s \quad , \\ 0 & \text{for } i^{*s} \in I_2^s \quad , \end{cases}$$

(3) $i^{*s} \notin I_0^s$,

$$(4) \quad w_i^{s+1} := \begin{cases} w_i^s - r^{s+1} c_i^s & \text{for } i \in I \setminus \{i^{*s}\}, \\ r^{s+1} v^s & \text{for } i = i^{*s}. \end{cases}$$

There are many possibilities to satisfy conditions (1) to (4). For example if $\{i \in I_2^s \mid c_i^s \neq 0\} \neq \emptyset$:

select $i^* := i^{*s} \in I_2^s$ with $c_{i^*}^s \neq 0$ arbitrarily. Let

$\{i_1, \ldots, i_k\} := \{i \in I_1^{s+1} \mid w_i^{s+1} = 0\}$, w_i^{s+1} defined by (4) ($i_j < i_{j'}$ for $j < j'$),

Then set $u_p^{s+1} := y_{i_p}^{s+1}$ for $1 \leq p \leq k$ and complete $(u_p^{s+1})_{1 \leq p \leq k}$ to a basis of $\mathbb{R}^n$ by proper elements of $\mathbb{R}^n$.

If $\{i \in I_2^s \mid c_i^s \neq 0\} = \emptyset$:

let $p^* := p^{*s} := \max \{p \in I \cup \{0\} \mid \bigvee i \in J_p^s \quad c_i^s > 0\}$

Select $i^* = i^{*s} \in J_{p^*}^s$ with $c_{i^*}^s > 0$ and

$$\frac{w_{i^*,p^*}^s}{c_{i^*}^s} := \min \left\{ \frac{w_{i,p^*}^s}{c_i^s} \;\middle|\; i \in J_{p^*}^s, \; c_i^s > 0 \right\}.$$

Let $\{i_1, \ldots, i_k\} := \{i \in \bigcup_{p^* \leq p \leq n} J_p^s \mid w_{i,p^*}^{s+1} = 0\}$.

Let $u_p^{s+1} := u_p^s$ for $1 \leq p \leq p^{*s}$, $u_{p^*+p}^{s+1} := y_{i_p}^{s+1}$ for $1 \leq p \leq k$ and complete $(u_p^{s+1})_{1 \leq p \leq p^*+k}$ to a basis of $\mathbb{R}^n$.

In any case conditions (1) to (4) are satisfied.

To start the algorithm we need a starting point $\overset{-1}{w} := (\overset{-1}{w}_t)_{t \in T}$. According to the construction of the algorithm we can start with a degenerate feasible basic solution.

Let $a_o \in \mathbb{R}$ such that $a_o < \bar{a} \leq \underline{a} = \inf \{< u, x > \mid x \in M\}$. Then: $\bigwedge x \in M, < x, u > \geq a_o$. Let $o \in T_1$ and $y_o := u$, $b_o := a_o$. Then

$$\overset{-1}{w}_t := \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{else} \end{cases}$$

is a feasible (degenerate) basic sulution for (D) and we

start the algorithm with $\bar{w}^1 := (\bar{w}^1_t)$.

THEOREM 6. (Convergence):

Let (V) be satisfied, R bounded and

(i) $\bigvee \gamma \in (0,1) \bigwedge s \in \mathbb{N} \quad d^s \geq \gamma g(x^s) > 0$ or

(ii) $\bigvee \{\eta\}_{n \in \mathbb{N}} \subset \mathbb{R}^+, \ \lim_{s \to \infty} \eta^s = 0$:

$$d^s \geq g(x^s) - \eta^s \ \text{and} \ d^s > 0.$$

Then the algorithm is convergent.

The proof is very long and complicated and will be published elsewhere.

4 Applications

a. Parabolic boundary control problem

The problem of optimal heating of a metal in a furnace leads to the following optimal control problem (see K. Glashoff and W. Krabs [2]). Given $z \in C[0,1]$, find a control $u \in L_\infty[0,T]$, $0 \leq u(t) \leq 1$ a. e., and a sulution y of the boundary value problem

$$\frac{\partial}{\partial t} y(x,t) - \frac{\partial^2}{\partial x^2} y(x,t) = 0, \ (x,t) \in (0,1) \times (0,t),$$

$$\alpha \frac{\partial}{\partial x} y(1,t) + y(1,t) = u(t), \ 0 < t < T,$$

$$\frac{\partial}{\partial x} y(0,t) = 0 \qquad , \ 0 < t < T,$$

$$y(x,0) = 0 \qquad , \ 0 < x < 1,$$

such that

$$\max_{0 \leq x \leq 1} |z(x) - y(x,T)|$$

has minimal value $\underline{a}$. This problem has a solution and in the case $\underline{a} > 0$, the corresponding optimal control u is bang-bang with a finite number of switching points in $[0,T-\varepsilon]$ for each $\varepsilon > 0$. Therefore we partition the interval $[0,T]$ into subintervals of equal length and consider control functions

u, constant on each subinterval. The remaining finite dimensional Tschebyscheff approximation problem under side conditions is solved with the algorithm in the case $\alpha = 0,1$, $z(x) \equiv 0,2$ and $T = 0,2$, $T = 0,3$.

b. <u>Hyperbolic boundary control problem</u>

We consider a vibrating string of length one, kept fixed at the left hand side and controlled by a force at the right hand side. The string has a given initial state and we intend to control it such that the whole energy at a fixed time T is minimal. This problem is described by the following mathematical model:

Find a control function

$$v \in V := \{v \in C^1 A\, [0,T] \;|\, v(0) = v'(0) = 0,\ \|v''\|_\infty \leq 1\}$$

and a solution $y(x,t;v)$ of the boundary-initial value problem

$$\frac{\partial^2}{\partial t^2} y(x,t) - \frac{\partial^2}{\partial x^2} y(x,t) = 0, \quad (x,t) \in (0,1) \times (0,T),$$

$$\left.\begin{aligned} y(x,0) &= y_0(x) \\ \frac{\partial}{\partial t} y(x,0) &= y_1(x) \end{aligned}\right\} \quad x \in [0,1],$$

$$\left.\begin{aligned} y(0,t) &= 0 \\ y(1,t) &= v(t) \end{aligned}\right\} \quad t \in [0,T],$$

such that $\int_0^1 \left[\left(\frac{\partial}{\partial t} y(x,T;v)\right)^2 + \left(\frac{\partial}{\partial x} y(x,T;v)\right)^2\right] dx$

has minimal value $\underline{a}$.

When y_0, y_1 are proper functions, this problem has a solution with $\underline{a} \neq 0$ for $T < 2$ which satisfies a weak bang-bang principle (see W. Krabs [3]). One can solve the differential equation problem explicitly. Then we have to minimize

$\int_{-1}^{+1} G'_v\,(T + x)^2 dx$ over $v \in V$ and

$$G'_v(T+x) := \begin{cases} G'(T+x) & \text{for } T+x \le 1, \\ G'(T+x-2) + v'(T+x-1) & \text{for } T+x \ge 1, \end{cases}$$

$$G(x) := \frac{1}{2}\left(y_o(x) + \int_o^x y_1(s)ds\right).$$

For computational purposes we partition the interval $[0,T]$ into subintervals and select a subspace of V with a basis consisting of suitable hat functions. Then we solve a quadratic program with a finite number of side conditions. Numerical results will be presented elsewhere.

References

1 Carasso, C. and P.J. Laurent, Un algorithme pour la minimisation d'une fonctionelle convex sur variete affine, Preprint Genoble 1973.

2 Glashoff, K. and W. Krabs, Dualität und Bang-Bang-Prinzip bei einem parabolischen Rand-Kontrollproblem, To appear in "Bonner Mathematische Schriften", 1975.

3 Krabs, W., Ein Kontroll-Approximationsproblem für die schwingende Saite, to appear in ISNM, 1975.

4 Schäfer, E., Ein Konstruktionsverfahren bei allgemeiner linearer Approximation, Numer. Math. 18 (1971), 113-126.

K.-H. Hoffmann and A. Klostermaier
III. Mathematisches Institut
der Freien Universität Berlin
Arnimallee 2-6
1000 Berlin 33

M-IDEALS IN APPROXIMATION THEORY

R. B. Holmes

This paper discusses some approximation theoretic properties and applications of M-ideals in Banach spaces. Several open questions are raised.

Let M be a closed linear subspace of the Banach space X. M is a <u>summand</u> of X (resp. an <u>ideal</u> in X) if it (resp. its annihilator M°) has a complementary subspace in X (resp. in X^*). All summands are clearly ideals, but the converse is false.

We are going to consider a special kind of ideal. Recall that an L-projection (resp. an M-projection) on X is a (bounded linear) projection P satisfying $\|x\| = \|Px\| + \|x - Px\|$ (resp. $\|x\| = \max(\|Px\|, \|x - Px\|)$), for $x \in X$. Then an L-<u>summand</u> (resp. an M-<u>summand</u>) is the range of an L-projection (resp. of an M-projection).

LEMMA 1. <u>The following properties of the subspace</u> M <u>are equivalent</u>:

M^o <u>is an</u> L-<u>summand of</u> X^*;

M^{oo} <u>is an</u> M-<u>summand of</u> X^{**};

<u>there is a projection</u> $P:X^* \to M^o$ <u>such that the unit ball</u> U^* <u>splits</u>: $U^* = \overline{co}(P(U^*) \cup (I - P)(U^*))$.

The proof can be read out of [1,2]. A subspace of X with any and hence all of these properties is called an M-<u>ideal</u> in X. Using the natural duality between L- and M-projections, it is easy to see that every M-summand is an M-ideal, and that the converse is true in reflexive spaces X. Since any L-projection is obviously a metric projection, the

Preceding page blank

annihilator of an M-ideal must be a Chebyshev subspace. Consequently, by a well-known theorem of Phelps, an M-ideal has the unique Hahn-Banach extention property.

A variety of examples of particular M-ideals is given in [1]. Of special interest is the fact [4] that the space of compact operators on certain Banach spaces X (e.g., c_o, ℓ^p, $1 < p < \infty$) is an M-ideal in the corresponding space of bounded operators. The characterization of such spaces in general is an open question.

LEMMA 2. *Let* M *be an* M-*ideal in* X. *Then for any* $\rho > 0$

$$(1) \quad (1 + \rho)U \cap \overline{(U + M)} \subset \overline{U + \rho U(M)}.$$

Proof. (Adapted from [2]). Let x belong to the left hand side of (1), and let $Q: X^{**} \downarrow M^{oo}$ be an M-projection. Then $Qx \in (1 + \rho)U(M^{oo})$. There hence exists $w \in \rho U(M^{oo})$ such that $Q(x - w) = Qx - w \in U^{**}$. Then $(I - Q)(x - w) = (I - Q)x$ also belongs to U^{**}, whence $x - w \in U^{**}$. That is,
$x = (x - w) + w \in (U^{**} + \rho(U(M^{oo})) \cap X \subset \overline{U + \rho U(M)}^{*} \cap X$
$= \overline{U + \rho U(M)}$, thus proving (1).

COROLLARY 1. *An* M-*ideal is proximinal*.

Proof. For any closed subspace M of X let Q_M be the quotient map. Then M is proximinal iff $Q_M(U)$ is closed in X/M. Now suppose that M is an M-ideal in X and that $Q_M(x) \in \overline{Q_M(U)}$. We shall find $y \in U$ so that $Q_M(x) = Q_M(y)$. Now $x \in \overline{U + M}$ so by (1) $x \in \overline{U + \rho U(M)}$ for any $\rho \geq \|x\| - 1$. Hence there exists $x_1 \in \rho U(M)$ such that $\|x - x_1\| < 1 + 1/2$. Then $x - x_1 \in \overline{U + 1/2U(M)}$ and so there exists $x_2 \in 1/2U(M)$ such that $\|x - x_1 - x_2\| < 1 + 1/4$. Continuing inductively we produce a sequence $\{x_n\}_2^\infty$ with $x_n \in 2^{1-n}U(M)$ and $\|x - x_1 - \ldots - x_n\| < 1 + 2^{-n}$. Therefore, if we set

$y = x + \sum_{1}^{\infty} x_n$, we see that $\|y\| \leq 1$ and $y - x \in M$, qed.

This result was proved by different methods in [1]. In [2] it was shown that this argument can be refined to further yield $\|y - x\| \leq \|x\| - 1 + \varepsilon$, for any given $\varepsilon > 0$. From now on we let P_M be the set-valued metric projection on M. In [7] the following result on simultaneous approximation was established.

LEMMA 3. *Let M be an M-ideal in X and let K be a compact subset of $X \setminus M$ such that $0 < \operatorname{diam}(K) < \varepsilon < \operatorname{dist}(K,M)$. Then given $y \in K$ and a best approximation $z \in P_M(y)$, there exists $\overline{z} \in \cap\{P_M(x) : x \in K\}$ with $\|z - \overline{z}\| < 2\varepsilon$.*

This result leads to the Lipschitz inequality

$$(2) \quad d_H(P_M(x), P_M(y)) \leq 2\|x - y\|, \quad x,y \in X.$$

Here, d_H is the Hausdorff metric on the nonempty closed bounded subsets of M, and the estimate (2) is sharp [7].

We now consider a second and perhaps more surprising consequence of Lemma 3. We say for brevity that a proximinal subspace M is *anti-Chebyshev* if

$$(3) \quad \operatorname{span} P_M(x) = M, \quad x \in X \setminus M.$$

THEOREM 1. *An M-ideal is anti-Chebyshev.*

Proof. By Corollary 1 any M-ideal M is proximinal. Now for a given $x \in X \setminus M$ let $V = \operatorname{span} P_M(x)$ and select any $y \in M$. Apply Lemma 3 to the set $\{x, x + \alpha y\}$ for sufficiently small $\alpha > 0$. If $w \in P_M(x) \cap P_M(x + \alpha y)$, then $w = \alpha y + z$ for some $z \in P_M(x)$. Solving for y then shows $y \in V$, q.e.d.

See [7] for further consequences and discussion of property (3).

If M is an ideal in X there is an operator $T:X \to X^{**}$ such that $T|M$ is the identity injection and $T(X) \subset M^{oo}$. If M is actually an M-ideal then Lemma 1 can be employed to show that T also satisfies

$$(4) \quad \|x\| = \max(\operatorname{dist}(x,M), \|Tx\|), \quad x \in X.$$

If M^θ is the metric complement of M it follows from (4) that T is norm preserving on the open set $X \setminus M^\theta$ and that M^θ contains the subspace $\ker(T)$. Of course, M^θ is never itself a subspace on account of Theorem 1.

If M is an M-ideal then M^θ can fail to have nonempty iterior [7], but it will if M is an M-summand. Does this property in fact characterize M-summands among M-ideals?

Examination of the proof of Theorem 1 and that of (2) in [7] shows that both these approximation properties of M-ideals depend only on the "2-ball property" of [1]. Now the 2-ball property is not characteristic of M-ideals and as far as we know this class of subspaces has not been satisfactorily described. Must a subspace with the 2-ball property necessarily be an ideal?

Property (2) is known to hold for other subspaces besides M-ideals. For example, it is valid for the subspace of continuous functions in the (real) space of bounded Borel functions on a paracompact topological space [5, p. 173]. However, although an ideal, this subspace lacks the 2-ball property. In contrast, this subspace need not satisfy property (3). In fact, it can happen that some bounded functions have unique best continuous approximants. From this evidence our anti-Chebyshev property appears very restrictive.

M-ideal theory and techniques provide a powerful and unified approach to compact operator approximation. To illustrate this assertion we mention a recent result due to Chui et al. [3]. Let H be a Hilbert space and let $B(H)$ (resp. $C(H)$) be the algebra of all bounded (resp. compact) operators on H.

THEOREM 2. For all $T \in B(H)$ we have $\cap\{P_{C(H)}(T + \lambda I): \lambda \text{ scalar}\} \neq \emptyset$.

Thus, there is a compact operator of simultaneous best approximation to the line $\{T + \lambda I\}$. It follows that given $T \in B(H)$ there exists $K \in C(H)$ such that $\|\rho(T + K)\| = \|\rho(T)\|_e \equiv \text{dist}(\rho(T), C(H))$, for all linear polynomials ρ. Whether such a K can be found for higher degree polynomials is an open question.

As has also been observed in [10], this compact operator K has a further remarkable property. To state it, let $W_o(T)$ be the algebra numerical range of T [9], and let $W_e(T) \equiv W_o(T + C(H))$ be the essential numerical range of T.

COROLLARY 2. For each $T \in B(H)$ there exists $K \in C(H)$ satisfying $W_e(T) = W_o(T + K)$.

Proof. Since $W_e(T) = \cap\{W_o(T + K): K \in C(H))$, we certainly have inclusion from left to right. Now if K is the compact operator of Theorem 2, and if $t \in W_o(T + K)$, then $|t - \lambda| \leq \|T + K - \lambda I\| = \|T - \lambda I\|_e$ for all λ, proving that $t \in W_e(T)$, q.e.d.

This corollary derives some of its interest by contrast with the recent result $\sigma_e(T) = \sigma(T + K)$ for $T \in B(H)$ [8]; here $\sigma_e(T) \equiv \cap\{\sigma(T + K): K \in C(H)\}$ is the Weyl essential spectrum of T.

It is not clear what operators other than I can serve in Theorem 2. We are in effect asking what lines belong to

the metric complement $C(H)^{\theta}$. Since the contents of $C(H)^{\theta}$ are only incompletely known [6] this will be a difficult question to answer in any generality.

References

1 Alfsen, E. and E. Effros, Structure in real Banach spaces, Amer. Math. 96 (1972), 98-173.

2 Ando, T., Closed range theorems for convex sets and linear liftings, Pac. J. Math. 44 (1973), 393-410.

3 Chui, C., D. Legg, P. Smith and J. Ward, On a question of Olsen concerning compact perturbations of operators, preprint.

4 Hennefeld, J., A decomposition for B(X)* and unique Hahn-Banach extensions, Pac. J. Math. 46 (1973), 197-199.

5 Holmes, R., A Course on Optimization and Best Approximation, Springer-Verlag, 1972.

6 Holmes, R., B. Scranton and J. Ward, Best approximation by compact operators II, Bull. AMS 80 (1974), 98-102.

7 Holmes, R., B. Scranton and J. Ward, Approximation from the space of compact operators and other M-ideals, Duke Math. J. 42 (1975), 259-269.

8 Stampfli, J., Compact perturbations, normal eigenvalues, and a problem of Salinas, J. London Math. Soc. 9 (1974) 165-175.

9 Stampfli, J. and J. Williams, Growth conditions and the numerical range in a Banach algebra, Tôhoku Math. J. 20 (1968), 417-424.

10 Ward, J., C. Chui, P. Smith, and R. Smith, L-ideals and numerical range preservation, preprint.

R. B. Holmes
Department of Mathematics
Purdue University
W. Lafayette, IN 47907

ROLLE REGULAR BIRKHOFF INTERPOLATION

L. Jaffe

1 Rolle regularity

In this paper we discuss problems of regularity of Birkhoff interpolation matrices, and introduce the new notion of Rolle regularity. The matrix $E = (e_{ik})_{i=1}^{m}{}_{k=0}^{n}$ has elements e_{ik} that are zeros or ones, and has exactly $n + 1$ ones. Let $[a,b]$ be a given interval, let X: $a \leq x_1 < \ldots < x_m \leq b$ be a set of knots in $[a,b]$. The pair E, X is regular if the interpolation problems

$$(1) \quad P^{(k)}(x_i) = c_{ik}, \; e_{ik} = 1,$$

have unique solutions in the space of polynomials P of degree $\leq n$. The matrix E is regular, if E, X is regular for each X. In what follows, we use the terminology of [2], [4], [6] without explanation.

DEFINITION 1. A pair E, X is Rolle regular if for each $f \in C^n[a,b]$, conditions

$$(2) \quad f^{(k)}(x_i) = 0, \; e_{ik} = 1,$$

imply the existence of a ξ, $a < \xi < b$ with the property

$$(3) \quad f^{(n)}(\xi) = 0.$$

The matrix E is Rolle regular, if E, X is Rolle regular for each X.

A matrix E is conservative (Birkhoff [1]) if it has no odd supported sequences.

THEOREM 1. A conservative matrix which satisfies the Birkhoff (strong Pólya) condition is Rolle regular.

This follows from the usual proofs of the Atkinson-Sharma theorem (see [4], [5]), which give also the fact (not mentioned there) that the possibilities $\xi = a, b$ can be eliminated.

Of great importance in Birkhoff interpolation is the determinant $D_E(X)$ of the system (1), and the Birkhoff kernel $K_E(X,t)$, $a \leq t \leq b$ (see [1], [3], [4]). We recall the following facts:

(4) $\int_a^b K_E(X,t)dt = D_E(x)$;

(5) If $f \in C^n[a,b]$ satisfies (2), then $\int_a^b f^{(n)}Kdt = 0$;

If $g \in C[a,b]$ has the property $\int_a^b gKdt = 0$, then there

(6) exists a function $f \in C^n[a,b]$ which satisfies (2) and

has $f^{(n)} = g$.

DEFINITION 2. (Birkhoff [1]). An interpolation matrix E is simple, if for each set of knots X, the kernel $K_E(X,t)$ is of constant sign, and does not vanish identically.

In this case, $K(X,t)$ is of the same sign for all X,t. Indeed, let $K(X',t') > 0$, $K(X'',t'') < 0$. We may assume that all sets of knots X in question have the same first and last elements $x_1 < x_m$. For each $X, K(X,t)$ is a continuous not identically vanishing function on $[x_1,x_m]$, which has a constant sign, and it is positive if $X = X'$, negative if $X = X''$. Changing X continuously from X' to X'', we obtain

a contradiction.

THEOREM 2. *Let* E *be a regular matrix*. (i) *For a given set of knots* X, *the pair* E,X *is Rolle regular if and only if* $K_E(X,t)$ *is of constant sign*. (ii) E *is Rolle regular if and only if it is simple*.

Proof. If K is of constant sign, and if $f \in C^n[a,b]$ satisfies (2), then from (4) and (5),

$$\int_a^b f^{(n)} K dt = 0, \quad \int_a^b K dt = D_E(X) \neq 0.$$

We deduce that $f^{(n)}$ cannot be strictly positive or strictly negative on [a,b], hence it satisfies (3).

If K changes sign for some X, then there is a strictly positive continuous function g with $\int_a^b gKdt = 0$. For f defined by (6), we have (2), but (3) does not hold. This proves (i).

One can sometimes reduce the requirements of Rolle regularity to some subclasses of $C^{(n)}$.

THEOREM 3. (L. Jaffe and G. G. Lorentz) *Let* E *be a regular matrix*. *Assume that no polynomial* P (*of any degree*) *with the property*

$$(7) \quad P^{(n)}(x) = (x - \alpha)^p (\beta - x)^q, \ \alpha < a, \ b < \beta, \ p, \ q = 0,1,\ldots$$

can be annihilated by E,X. *Then* E,X *is Rolle regular.*

Proof. For each selection $\alpha < a$, $\beta > b$, p, $q = 0,1,\ldots$ we must have $F_{pq}(\alpha,\beta) \neq 0$, where

$$(8) \quad F_{pq}(\alpha,\beta) = \int_a^b (t - \alpha)^p (\beta - t)^q K(t) dt.$$

For otherwise (6) would give us a polynomial P satisfying (7) and annihilated by E,X.

Without loss of generality let $D_E(X) > 0$. Since

$$\left(\frac{1}{-\alpha}\right)^p \left(\frac{1}{\beta}\right)^q F_{pq}(\alpha,\beta) = \int_a^b (1 - \frac{t}{\alpha})^q K dt \to D_E(X) > 0$$

for $\alpha \to -\infty$, $\beta \to +\infty$, and since $F_{pq}(\alpha,\beta)$ never vanishes, it follows that $F_{pq}(\alpha,\beta) > 0$ for all choices of parameters.

If $c_k \geq 0$, $k = 0,\ldots,N$, we derive from this

$$\int_a^b \sum_{k=0}^{N} c_k (t - \alpha)^k (\beta - t)^{N-k} K(t)dt \geq 0.$$

If g is positive continuous function on $[\alpha,\beta]$, it is possible to select the constants c_k so that the sum under the integral sign will become the Bernstein polynomial $B_N(g)$ of g on $[\alpha,\beta]$. For $N \to \infty$, the relation $\int_a^b B_N(g)Kdt \geq 0$ becomes $\int_a^b gKdt \geq 0$. Since g was arbitrary, we must have $K(t) \geq 0$ for all t in $[a,b]$.

In terms of Rolle regularity of matrices, this result has the following formulation.

THEOREM 3A. <u>A regular</u> $m \times (n + 1)$ <u>matrix</u> E <u>with</u> $m \geq 2$, <u>with</u> $n + 1$ <u>ones and with zeros in the last column is strongly regular if and only if each matrix</u> E' <u>derived from</u> E <u>by adding to it rows</u> 0 <u>and</u> $m + 1$ <u>with ones in positions</u> $(0,n)$, $(0,n+1)$, $\ldots$, $(0,n+p-1)$; $(m+1,n)$, $\ldots$, $(m+1, n+q-1)$, $p,q = 0,1,\ldots$ <u>is regular</u>.

Regular must be also the matrices obtained from E' by coalescing row 0 to row 1, or row $m + 1$ to row m (or both).

We note two more results, which can be derived from the above. For a matrix E, let $\overline{E}$ (or $\underline{E}$) be the $(m - 1) \times n$ matrix derived from E by omitting from it the last column and the last one of the first (or the last) row.

PROPOSITION 1. <u>A Birkhoff matrix</u> E <u>cannot be Rolle regular if one of the matrices</u> $\overline{E}$, $\underline{E}$ <u>is strongly singular</u>.

PROPOSITION 2. Let E be a Birkhoff matrix with row 1 or row m containing at least two ones, or let E be a three row Birkhoff matrix. If one of the rows of E contains exactly two odd supported sequences, one of them terminating in the penultimate column, then $K(X,t)$ changes sign.

2 Counterexamples

For our main properties we have established the following implications:

$$\text{CONSERVATIVE} \Longrightarrow \text{ROLLE REGULAR} \Longleftrightarrow \text{SIMPLE} \Longrightarrow \text{REGULAR}.$$

The first and the last implications cannot be inverted:

EXAMPLE 1. The matrix

$$E = \begin{Vmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{Vmatrix}$$

is regular, but not Rolle regular.

The regularity of E has been noted in [5]. We take $X = (0, x, 1)$; the kernel $K_E(X,t)$ changes sign for $0 \le x \le t \le 1$. This follows from the formula

$$K(X,t) = -\frac{x(1-t)^3}{6 \cdot 4!}\{3x(1+t) - 1 - 3t\}.$$

(This result follows also from Proposition 2).

EXAMPLE 2. The matrix

$$E = \left\| \begin{matrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right\|$$

is Rolle regular, but not conservative.

This Birkhoff matrix has two odd supported sequences and is not conservative. We show that it is Rolle regular by explicit computation of its kernel $K_E(X,t)$. The proof that $K(X,t) \leq 0$ for all X,t is elementary, but not quite simple. Without loss of generality we can take $x_1 = 0$, $x_2 = x$, $x_3 = 1$. From the determinant expression for $K(X,t)$ we obtain

$$K_E(X,t) = \begin{vmatrix} \frac{1}{4!}(x-t)_+^4 & \frac{1}{4!}x^4 & \frac{1}{3!}x^3 \\ \frac{1}{2}(x-t)_+^2 & \frac{1}{2}x^2 & x \\ \frac{1}{5!}(1-t)_+^5 & \frac{1}{5!} & \frac{1}{4!} \end{vmatrix}$$

Hence

$$K_E(X,t) = \begin{cases} -\frac{1}{4!}\frac{1}{5!}(1-t)^5 x^5 & \text{for } x \leq t \leq 1 \\ -\frac{xt^2}{2\cdot 4!\,5!}\, f(x,t) & \text{for } 0 \leq t \leq x, \end{cases}$$

where

$$f(x,t) = (20x^2 - 25x + 8)x^2 - 4tx(5x^3 - 5x + 2) + t^2(10x^4 - 5x + 2) - 2t^3x^4.$$

Evidently $K_E(X) \neq 0$ and also $K_E(X,t) \leq 0$ for $x \leq t \leq 1$. To show that this inequality holds also for $0 \leq t \leq x$, we use critically the following equivalent expression for $f(x,t)$:

$$f(x,t) = 2t^2(1 - x)^5 + 2t(x - t)(1 - x)^4(x + 4)$$

$$+ (x - t)^2(-2x^4t + 20x^2 - 25x + 8).$$

If $0 \leq t \leq x/2$ we have

$$f(x,t) \geq (x - t)^2(-x^5 + 20x^2 - 25x + 8).$$

Let $g(x) = -x^5 + 20x^2 - 25x + 8$. In $0 \leq x \leq 1$ the function g attains its minimum inside the interval at a point x where $x^4 = 8x - 5$ and $0 < x < 1$. Hence at the minimum

$$g(x) = 12x^2 - 20x + 8 = 4(1 - x)(2 - 3x).$$

But at the minimum $x = 5/8 + (1/8)x^4$. Hence $x < 5/8 + 1/8 = 3/4$ which in turn gives the better estimate $x < 5/8 + 1/8(3/4)^4 < 2/3$, showing that $g(x) \geq 0$ for $0 \leq x \leq 1$.

On the other hand if $(1/2)x \leq t \leq x$ we have $t \geq x - t$. Hence

$$f(x,t) \geq (x - t)^2[2(1 - x)^5 + 2(1 - x)^4(x + t)$$

$$-2x^4t + 20x^2 - 25x + 8]$$

and it suffices to show that

$$h(x) = 10(1 - x)^4 - 2x^5 + 20x^2 - 25x + 8$$

$$= 2(1 - x)^5 - 20x^3 + 60x^2 - 55x + 16 \geq 0$$

for $0 \leq x \leq 1$. In turn it suffices to show that

$$k(x) = -20x^3 + 60x^2 - 55x + 16 \geq 0 \quad \text{for } 0 \leq x \leq 1.$$

The function k is at a minimum where $x^2 = 2x - 11/12$ at which point $k(x) = (10/3)(x - 7/10)$. The minimum occurs at $x = 1 - \sqrt{1/12} = 0.712$ which is greater than $7/10$!

This completes the proof that $K_E(X)$ does not change sign and hence that E is Rolle regular.

References

1 Birkhoff, G. D., General mean value and remainder theorems with applications to mechanical differentiation and integration, Trans. Amer. Math. Soc., 7 (1906), 107-136.

2 Lorentz, G. G., The Birkhoff interpolation problem: New methods and results, in Proceedings Int. Conference, Oberwolfach, Birkhauser Verlag, Basel 1974, pp. 481-501.

3 Lorentz, G. G., Zeros of splines and Birkhoff's kernel, Math. Z. 142 (1975), 173-180.

4 Lorentz, G. G., Birkhoff interpolation problem, Report CNA-103, Center of Numerical Analysis, The University of Texas, July 1975, 117 pp.

5 Lorentz, G. G., and K. L. Zeller, Birkhoff interpolation, SIAM J. Numer. Anal. 8 (1971) 43-48.

6 Lorentz, G. G., and K. L. Zeller, Birkhoff interpolation problem: Coalescence of rows. Arch. Math. (Basel) 26 (1975), 189-192.

L. Jaffe
Department of Mathematics
The University of Texas
Austin, Texas 78712

BIRKHOFF INTERPOLATION BY SPLINES

K. Jetter

This paper deals with regularity theorems concerning the problem of Birkhoff interpolation by splines. In particular, it is shown how the Pólya condition can be extended to spline interpolation. A duality theorem and a generalized Atkinson and Sharma theorem are treated in detail.

1. Introduction

Let $k \geq 0$, $m \geq 1$, $n \geq 0$ be fixed natural numbers, $X := \{-1 = x_0 < x_1 < \ldots < x_{n+1} = +1\}$ and $X^* := \{-1 = x_0^* < x_1^* < \ldots < x^*_{k+1} = +1\}$ fixed real point sets and

$$E = (e_{i,j})_{i=0}^{n+1}{}_{j=0}^{m-1}, \quad E^* = (e^*_{p,q})_{p=1}^{k}{}_{q=0}^{m-1}$$

fixed incidence matrices (i.e., $e_{i,j}$, $e^*_{p,q} = 0$ or $= 1$). The problem of Birkhoff interpolation by splines consists in finding conditions on the regularity of (E,X,E^*,X^*): Does there exist, for every real data $c_{i,j}$, a unique spline $s \in S$,

$$S := \langle 1,\ldots,t^{m-1}, \underbrace{\ldots,(t-x_p^*)_+^{m-1-q},\ldots}_{e^*_{p,q}=1} \rangle,$$

solving the equations $s^{(j)}(x_i) = c_{i,j}$ for $e_{i,j} = 1$?

It is natural to assume that the number $|E|$ of interpolation conditions is equal to the dimension of S,

(A1) $|E| = m + |E^*|$,

and in order to have well-defined derivatives $s^{(j)}(x_i)$ for $e_{i,j} = 1$, we require that

(A2) $x_i = x_p^* \Rightarrow e_{i,j} + e^*_{p,m-1-j} \leq 1, \quad j = 0,\ldots,m-1.$

2. The generalized Pólya condition

In the special case of polynomial interpolation (where $|E^*| = 0$) the Pólya condition is well known (see Schoenberg [7]). In a possible formulation this condition requires that no subblock of interpolation conditions may produce an overdetermined interpolation problem. Consider a triple (r,i_1,i_2), $0 \leq r \leq m-1$, $0 \leq i_1 \leq i_2 \leq n+1$, and let

$$E = \left[\begin{array}{c} \boxed{E(r,i_1,i_2)} \\ \underset{r}{\vdots} \end{array}\right] \begin{array}{l} -i_1 \\ -i_2 \end{array}, \qquad E^* = \left[\begin{array}{c} \boxed{E^*(r,i_1,i_2)} \\ \underset{m-1-r}{\vdots} \end{array}\right] \begin{array}{l} -p_1 \\ -p_2 \end{array}$$

where $x^*_{p_1} \leq x_{i_1} < x^*_{p_1+1}$, $x^*_{p_2-1} < x_{i_2} \leq x^*_{p_2}$. With respect to the basis

$$1,\ldots,\frac{t^{m-1}}{(m-1)!},\underbrace{\ldots,\frac{(x^*_p-t)_+^{m-1-q}}{(m-1-q)!},\ldots}_{x^*_p \leq x^*_{p_1}},\underbrace{\ldots,\frac{(t-x^*_p)_+^{m-1-q}}{(m-1-q)!},\ldots}_{x^*_p > x^*_{p_1}}$$

of S the coefficient matrix associated with (E,X,E^*,X^*) takes the form $\Delta = \left[\begin{array}{c|c} \Delta_2 & \\ \hline 0 & \Delta_1 \end{array}\right]$ if rows and columns of Δ are ordered as follows: The lower rows are associated with interpolation conditions characterized by $e_{i,j} \in E(r,i_1,i_2)$, and the columns at the back belong to the basis functions $t^j/j!$, $r \leq j \leq m-1$, and $(t-x^*_p)_+^{m-1-q}/(m-1-q)!$ where $e^*_{p,q} \in E^*(r,i_1,i_2)$. If Δ_1 contains more rows than columns then Δ cannot be regular. We have shown (see Melkman [6] for another proof in the case of "Hermite" splines)

THEOREM 1. <u>The following generalized Pólya condition is</u>

<u>necessary for</u> (E,X,E^*,X^*) <u>to be regular</u>:

$$\text{(A3)}\quad \begin{array}{l} |E(r,i_1,i_2)| \le m - r + |E^*(r,i_1,i_2)| \\ \underline{\text{for all}}\ (r,i_1,i_2),\ 0 \le r \le m-1,\ 0 \le i_1 \le i_2 \le n+1. \end{array}$$

It is obvious, too, how D. Ferguson's [2] decomposition theorem carries over to spline interpolation. Call the problem (r,i_1,i_2)-decomposable, if $(r,i_1,i_2) \neq (0,0,n+1)$, and if equality holds in (A3). Then the matrices Δ_1 and Δ_2 are square and may be interpreted as the coefficient matrices of two easily identified interpolation problems (J_1) and (J_2).

THEOREM 2. <u>An</u> (r,i_1,i_2)-<u>decomposable problem</u> (E,X,E^*,X^*) <u>is regular if and only if the induced problems</u> (J_1) <u>and</u> (J_2) <u>are regular</u>.

3. Regular problems

Some regularity results can be proved using the properties of interpolation kernels. Recall that every regular problem (E,X,E^*,X^*) is characterized by its kernel $k(x,t) = \det \Delta^{-1} \det A(x,t)$,

$$A(x,t) = \left[\begin{array}{c|c} \dfrac{(x-t)_+^{m-1}}{(m-1)!} & 1 \ldots \dfrac{x^{m-1}}{(m-1)!} \ldots \dfrac{(x-x_p^*)_+^{m-1-q}}{(m-1-q)!} \ldots \\ \hline \vdots & \\ \dfrac{(x_i-t)_+^{m-1-j}}{(m-1-j)!} & \underbrace{\qquad\Delta\qquad}_{e^*_{p,q}=1} \\ \vdots & \end{array}\right] \Bigg\}\, e_{i,j}=1$$

Here Δ is the coefficient matrix of (E,X,E^*,X^*) with respect to the basis of S indicated by the first row.

Our first result is a "duality" theorem. We call $(\hat{E}^*,X^*,\hat{E},X)$ dual to (E,X,E^*,X^*) if $\hat{E}^*$ is obtained from E^* by adding

a new first row $(e^*_{0,0},\dots,e^*_{0,m-1})$ and a new last row $(e^*_{k+1,0},\dots,e^*_{k+1,m-1})$ where $e_{0,j} + e^*_{0,m-1-j} = e_{n+1,j} + e^*_{k+1,m-1-j} = 1$, $j = 0,\dots,m-1$, and if $\hat{E}$ is obtained from E by discarding its first and its last row.

THEOREM 3. $(\hat{E}^*,X^*,\hat{E},X)$ <u>is regular if and only if</u> (E,X,E^*,X^*) <u>is regular</u>.

Proof. Let $k(x,t)$ be the interpolation kernel associated with (E,X,E^*,X^*). Define $k^*(x,t)$ on the open square $(-1,+1)\times(-1,+1)$ by $k^*(x,t) := (-1)^m k(t,x)$. Then $k^*(x,t)$ is essentially the interpolation kernel associated with $(\hat{E}^*,X^*,\hat{E},X)$. For details see [3].

We shall now prove a generalization of the Atkinson and Sharma theorem (see [1]). For the concept of odd supported sequences consult Lorentz and Zeller [5].

THEOREM 4. <u>Assume that</u> (E,X,E^*,X^*) <u>satisfies</u> (A1), (A2), <u>and the generalized Pólya condition</u> (A3). <u>If</u> E <u>contains no odd supported sequence, and if</u> E^* <u>contains only sequences beginning in the first column, then</u> (E,X,E^*,X^*) <u>is regular</u>.

Melkman [6] gives a proof of this theorem in the case of simple spline nodes using a Budan-Fourier theorem for "Hermite" splines. His paper includes some hints for treating the general case.

The following proof using an induction on $|E|$ proceeds from a proof of the Atkinson and Sharma theorem due to Lorentz [4]. We may assume that $|E^*| > 0$ and $|E| > 2$.

If the problem (case I) is $(0,i_1,i_2)$-decomposable or (case II) is $(r,0,n+1)$-decomposable with $E^* = [E^*(r,0,n+1)\,|\,0\,]$, we can argue with the decomposition theorem, Theorem 2, using the induction hypothesis. Therefore assume that the cases I and II are excluded. Then it can be shown that there exist positions

$(i_0,0)$ and (p_0,q_0) with $e_{i_0,0} = e^*_{p_0,q_0} = 1$ such that the reduced problem $(\tilde{E},X,\tilde{E}^*,X^*)$ which is obtained from (E,X,E^*,X^*) by setting $e_{i_0,0} = e^*_{p_0,q_0} = 0$ is regular according to the induction hypothesis.

Consider the interpolation kernel $k(x,t)$ associated with $(\tilde{E},X,\tilde{E}^*,X^*)$, and put $k_0(t) := k(x_{i_0},t)$. Then (E,X,E^*,X^*) is regular if and only if $k_0^{(q_0)}(x^*_{p_0}) \neq 0$.

Kernels of this type have been studied in [3]. There is a minimal interval $[x_{i_1},x_{i_2}]$, $x_{i_1} \leq x_{i_0} \leq x_{i_2}$, in the exterior of which k_0 vanishes identically and in the interior of which k_0 vanishes only in discrete points. Using an estimation of zeros (see Lorentz [4], or [3]) k_0 has at most $|E(0,i_1,i_2)| - m - 1$ zeros, hence by (A3) at most $|E^*(0,i_1,i_2)| - 1$ zeros.

But if $e^*_{p,q} = 1$ and $(p,q) \neq (p_0,q_0)$, then $k_0^{(q)}(x^*_p) = 0$ as the determinant defining $k_0^{(q)}(x^*_p)$ has two columns differing at most in the sign. So, k_0 has exactly $|E^*(0,i_1,i_2)|-1$ zeros, and--as (p_0,q_0) must be at the end of a sequence--$x^*_{p_0}$ is contained in the open interval (x_{i_1},x_{i_2}). Moreover, we must have $k_0^{(q_0)}(x^*_{p_0}) \neq 0$. This completes the proof.

References

1 Atkinson, K., and A. Sharma, A partial characterization of poised Hermite-Birkhoff interpolation problems, SIAM J. Numer. Anal. 6 (1969), 230-235.

2 Ferguson, D., The question of uniqueness for G. D. Birkhoff interpolation problems, J. Approximation Theory 2 (1969), 1-28.

3 Jetter, K., Duale Hermite-Birkhoff Probleme, to appear in J. Approximation Theory.

4 Lorentz, G. G., Zeros of splines and Birkhoff's kernel, Math. Z. 142 (1975), 173-180.

5 Lorentz, G. G., and K. L. Zeller, Birkhoff interpolation, SIAM J. Numer. Anal. 8 (1971), 43-48.

6 Melkman, A. A., Hermite-Birkhoff interpolation by splines, manuscript 1975, 33 pp.

7 Schoenberg, I. J., On Hermite-Birkhoff interpolation, J. Math. Anal. Appl. 16 (1966), 538-543.

K. Jetter
Math. Inst. der Universität Tübingen
D-7400 Tübingen, BRD

NONUNIQUENESS OF SIMULTANEOUS APPROXIMATION

BY TRIGONOMETRIC POLYNOMIALS

Darell J. Johnson

We consider the approximation of differentiable functions by trigonometric polynomials of degree at most n, $\mathcal{T}_n$, in the norm

$$\|f\|_F \equiv \max_{i=0,\ldots,\rho}\|f^{(k_i)}\|,$$

where $0 = k_0 < k_1 < \ldots < k_\rho$ and $\|\cdot\|$ is the usual Chebyshev norm on $C(T)$. The result of interest obtained below is the determination of the set of best trigonometric polynomial approximations, in the sense that once one best approximate is known, all best approximates are known.

More specifically, let $F = \{k_0, k_1, \ldots, k_\rho\}$, where $0 = k_0 < k_1 < \ldots < k_\rho$ are integers. Let B_F be the Banach space of k_ρ-times continuously differentiable real-valued functions on the unit circle $T = [-\pi, \pi]$ under the F-norm defined as above.

By standard compactness arguments, given $f \in B_F$ there always exists at least one $q^* \in \mathcal{T}_n$ for which $\|f - q^*\|_F \leq \|f - p\|$ for all $q \in \mathcal{T}_n$. Thus, $\Omega(f) \equiv \{p \in \mathcal{T}_n;\ \|f - p\|_F \leq \|f - q\|_F$ for all $q \in \mathcal{T}_n\}$ is non-empty for all $f \in B_F$. Therefore $\Lambda(f) \equiv \|f - q^*\|, q^* \in \Omega(f)$, is well-defined, and positive if $f \notin \mathcal{T}_n$. Define extremal sets

$$U_i(p) \equiv \{x \in [-\pi,\pi); |(f-p)^{(k_i)}(x)| = \|f-p\|_F\}$$

LEMMA 1. *There exists a* $q^* \in \Omega(f)$ *having the property that* $U_i(q^*) \subset U_i(q)$ *and* $(q^*)^{(k_i)}(x) = q^{(k_i)}(x)$ *if* $x \in U_i(q^*)$, *for every* $q \in \Omega(f)$ *and every* $i = 0,\ldots,\rho$.

DEFINITION. *A* $q^* \in \Omega(f)$ *having the properties that* $U_i(q^*) \subset U_i(q)$ *and* q^* *agrees with* q *on* $U_i(q^*)$, *for every* $q \in \Omega(f)$ *and every* $i = 0,\ldots,\rho$, *is called a minimal polynomial of best approximation to* f.

To prove the lemma, note that each $U_i(q)$ is compact. Set $U_i = \bigcap\{U_i(q);\ q \in \Omega(f)\}$. Then there is a countable sequence $\{q_\nu\}_{\nu=1}^{\infty} \subset \Omega(f)$ for which $U_i = \bigcap_{\nu=1}^{\infty} U_i(q_\nu)$, for every $i = 0,\ldots,\rho$. Setting $q^* = \Sigma_{\nu=1}^{\infty} 2^{-\nu} q_\nu$, $q^* \in T_n$ by the uniform convergence of the series on the right and the fact that each $q_\nu \in T_n$. By convexity of $\Omega(f)$, each $q_\nu \in \Omega(f)$ also implies $q^* \in \Omega(f)$. If $q^{(k_j)} \in \Omega(f)$ differs from $(q^*)^{(k_j)}$ at some $x_0 \in U_j(q^*) = U_j$, then $U_j((q + q^*)/2)$ is a proper subset of $U_j(q^*)$, a contradiction.

As a corollary to our lemma, it is clear that if p^*, q^* are both minimal polynomials of best approximation to f, then $U_i(p^*) = U_i(q^*)$ and $p^*(x) = q^*(x)$ for all $x \in U_i(p^*) = U_i(q^*)$, for each $i = 0,\ldots,\rho$. Abusing notation slightly, let $U_i(f) \equiv U_i(q^*)$, q^* a minimal polynomial of best approximation to f, for each $i = 0,\ldots$. Necessarily $U_j \neq \emptyset$ for at least one $j \in \{0,\ldots,\rho\}$.

LEMMA 2. *A function* q^* *is a minimal polynomial of best approximation to* $f \in B_F \setminus T_n$ *if and only if there exists no* $q \in T_n$ *for which*

$$\max\{\sup[(f-q^*)^{(k_i)} q^{(k_i)}](x);\ x \in U_i(q^*);\ k_i \in F\} \leq 0$$

and such that the strict inequality

$$[(f - q^*)^{(k_j)} q^{(k_j)}](x_j) \leq 0$$

holds for some $j \in \{0,\ldots,\rho\}$ and some $x_j \in U_j(q^*)$.

The proof of this Kolmogoroff characterization theorem is standard (e.g., [3; p. 19]).

In particular, if q^* is a minimal polynomial of best approximation to f then $\Sigma_{j=0}^{\rho} |U_j(q^*)| \geq n + 1$, where $|U_j(q^*)|$ denotes the number of points in $U_j(q^*)$, for otherwise we may find a $q \in T_n$ that satisfies the interpolation data $q^{(k_j)}(x_\nu) = -(f - q^*)^{(k_j)}(x_\nu)$, for all points $x_\nu \in U_j(p^*)$, for all $j = 0,\ldots,\rho$.

LEMMA 3. If $f \in B_F \backslash T_n$ is $(k_i + 1)$-times differentiable on $[-\pi,\pi]$, then $p \in \Omega(f)$, $x \in U_i(p)$ implies $(f - p)^{(k_i+1)}(x) = 0$. Moreover $k_i + 1 = k_{i+1}$ implies that $U_i(p)$ and $U_{i+1}(p)$ are disjoint.

Proof. Assumption $x \in U_i(p)$ implies that $(f - p)^{(k_i)}$ has a maximum or a minimum at x, hence $(f - p)^{(k_i+1)}(x) = 0$. If $k_{i+1} = k_i + 1$, then $x \in U_i(p) \cap U_{i+1}(p)$ implies

$$\Lambda(f) = |(f - p)^{(k_{i+1})}(x)| = |(f - p)^{(k_i+1)}(x)| = 0,$$

a contradiction since $f \notin T_n$.

Set $G \equiv \{k_i \in F; U_i(f) \neq \emptyset\}$ and let $q \equiv \min\{k_i; k_i \in G\}$.

THEOREM. Suppose $f \in B_F \backslash T_n$ is $k_\rho + 1$ times differentiable. Then f has a unique best approximate in F-norm whenever $q = 0$, while $\Omega(f)$ is a convex set of dimension one if $q > 0$. Furthermore the derivative of any best approximate in F-norm is always unique.

Proof. Let $p^* \varepsilon \Omega(f)$ be a minimal polynomial of best approximation to f. Suppose $p \varepsilon \Omega(f)$. By Lemma 1, $p^{(k_i)}(x) = p^{*(k_i)}(x)$ whenever $x \varepsilon U_i(f)$, for each $k_i \varepsilon G$. Lemma 3 then implies $p^{(k_i+1)}(x) = p^{*(k_i+1)}(x)$ whenever $x \varepsilon U_i(f)$, $k_i \varepsilon G$. Setting $r \equiv p - p^*$, $r \varepsilon T_n$ then satisfies the constraints

$$(1) \quad r^{(k_i)}(x) = r^{(k_i+1)}(x) = 0$$

whenever $x \varepsilon U_i(f)$, $k_i \varepsilon G$. By Lemma 2, the $U_i(f)$, $k_i \varepsilon G$, must contain at least $n + 1$ points, counting the same point over again if it is contained in another $U_j(f)$. By Lemma 3, therefore, the constraints (1) number at least $2n + 2$, since $U_i(f) \cap U_{i+1}(f) = \emptyset$ if $k_{i+1} = k_i + 1$. Thus we may form a $(2n = 1)$-incidence matrix E' from the constraints (1) by choosing an $x_q \varepsilon U_q(f)$, imposing the constraint $r^{(q)}(x_q) = 0$ (while ignoring the constraint $r^{(q+1)}(x_q) = 0$), and then choosing n other pairs of constraints $r^{(k_i)}(x_\nu) = r^{(k_i+1)}(x_\nu) = 0$, $x_\nu \varepsilon U_i(f)$, $k_i \varepsilon G$. Form E from E' by deleting the first q columns of E'. The resulting $(2n + 1)$-incidence matrix E is strongly conservative and satisfies the weak Pólya condition, so by [2] is poised with respect to T_n. Since $s \equiv r^{(q)} \varepsilon T_n$ satisfies E, $s \equiv 0$, whence $r \varepsilon T_n$ must be a constant trigonometric polynomial. Moreover if $q = 0$, $r \equiv 0$, and $p = p^*$.

It is instructive to contrast the above result with the analogous algebraic polynomial result of R. A. Lorentz [3]. In the latter case the dimension of $\Omega(f)$ is always q, with the q'th derivative of a best approximate necessarily being unique. In the trigonometric case things are more civilized: the dimension of $\Omega(f)$ is at most one, for any value of q. This surprising result leads one to question whether a best trignometric approximation in F-norm is

necessarily unique--that is, can q ever be nonzero?

Example. Consider $f \in C^1(T)$ defined as follows:

$$f(x) = \begin{cases} x + \pi & , -\pi \leq x \leq -\pi/2 \\ (-2/\pi)x^2 - x + \pi/2 & , -\pi/2 \leq x \leq \pi/4 \\ (-10/\pi)x^2 - 5x & , -\pi/4 \leq x \leq 0 \\ -f(-x) & , 0 \leq x \leq \pi \ . \end{cases}$$

Considering approximating f' from $T_1^1 \equiv \langle \sin x, \cos x \rangle$. Since $f' \in C(T)$ is an even function, its (unique) best approximate in the usual Chebyshev norm must be an even trigonometric polynomial: $\alpha \cos x$, for some real α. By inspection $\alpha = -3$ yields the best approximate to f' from T_1^1, with deviation 2. On the other hand $q(x) = -3 \sin x$ has a maximum deviation of (approximately) $3 - \pi/2$, which is strictly less than 2. Thus $q = 1$ in this example. Note that any polynomial of the form $p(x) = -3 \sin x + c$, where $|c| < \varepsilon$, for $\varepsilon \simeq -1 + \pi/2$, is a best approximate to f here in the F-norm, where $F = \{0,1\}$.

References

1 Chalmers, B. L., Uniqueness of best approximation of a function and its derivatives, J. Approximation Theory, 7 (1973), 213-225.

2 Johnson, D. J., The trigonometric Hermite-Birkhoff interpolation problem, Trans. AMS, 212 (1975), 365-374.

3 Lorentz, R. A., Nonuniqueness of simultaneous approximation by algebraic polynomials, J. Approximation Theory 13 (1975), 17-23.

D. J. Johnson
New Mexico State University
Las Cruces, New Mexico

APPROXIMATION BY POLYNOMIALS WITH RESTRICTED COEFFICIENTS

D. Leviatan

The purpose of this paper is to give estimates on the rate of approximation of functions by means of polynomials with restricted coefficients. Also discussed is the possibility of approximating monotone functions by monotone polynomials of the above type and estimates on the rate of approximation of this type are obtained.

1 Introduction

Let $A = \{A_k\}(k \geq 1)$ be a sequence of non-negative constants and set

$$P_A = \{p : p(x) = \sum_{k=1}^{n} a_k x^k \qquad n = 1,2,\ldots, |a_k| \leq A_k^k\}$$

Denote by $C_o[0,1]$ the space of continuous functions in $[0,1]$ that vanish at the origin. Recently Roulier [6] and v. Golitschek [2] (see also [3]) have shown that in order for P_A to be dense in $C_o[0,1]$ it is necessary and sufficient that there should exist a subsequence $\{k_i\}$ of natural numbers with the properties that

$$(1) \quad \sum_{i=1}^{\infty} 1/k_i = \infty \text{ and } \lim_{i\to\infty} A_{k_i} = \infty$$

Since the rate of approximation of functions in $C_o[0,1]$ is known by Jackson's theorem, it is natural to ask how much, if any, is lost in the rate of approximation by restricting the polynomials to P_A for a prescribed A. In view of this Bak, v. Golitschek and the author [1] have called P_A efficient if the rate of approximation to functions in $C_o[0,1]$

Preceding page blank

by polynomials in P_A is still described by Jackson's estimate. This does not mean that a particular function is approximable at the same rate (see Example 1).

We will review some of the results in [1] in section 2 while in Section 3 we will discuss monotone approximation with elements in P_A.

2 Efficiency of P_A

We restrict ourselves here to the case where $A_k \neq 0$ for all $k \geq 1$ namely all powers are allowed. We have the following sufficient condition for efficiency (see [1] Thm. 2).

THEOREM 1. *Let* r *be a non negative integer and* $\delta > 0$. *Let* A *satisfy* $A_k \geq \delta k^2$, $k \geq k_o$. *Then there exists a constant* $K = K(r,\delta)$ *independent of* A *such that for each* $f \in C^r[0,1]$ *with* $f^{(i)}(0) = 0$ $i = 0,\dots,r$ *and any sufficiently large* n *there is an* n-th *degree polynomial* $p_n \in P_A$ *with*

$$(2) \quad ||f - p_n|| \leq Kn^{-r}\,\omega(f^{(r)};\tfrac{1}{n}).$$

Here and in the sequel $\omega(f;.)$ denotes the modulus of continuity of f. This condition is almost necessary in the sense that if $A_k = 0(k^{2-\varepsilon})$ for some $\varepsilon > 0$, then P_A is not efficient. We see this from the following (see [1] Thm 6).

THEOREM 2. *If* P_A *is efficient and if* $S_k = 0(k^{2-\varepsilon})$ *for some* $\varepsilon > 0$, *then there exists an infinite subsequence* k_i *such that* $A_{k_i} \geq S_{k_i}$, $i = 1,2,\dots$.

As was said in the introduction this is global efficiency and does not mean that a particular function is approximable at the same rate as obtained with the set of all polynomials. This is illustrated in the following.

EXAMPLE 1. Let $f(x) = x^{1/2}$. Then Jackson's theorem guarrantees that the rate of approximation to f by the set of all polynomials is at least of the order $n^{-1/2}$. In fact there exists an n-th degree polynomial p_n such that

$$(3) \qquad ||x^{1/2} - p_n(x)|| \leq \prod_{k=1}^{n} \frac{k - \frac{1}{2}}{k + \frac{1}{2}}$$

$$\leq \exp\left[-\sum_{k=1}^{n} 1/k\right]$$

$$\sim n^{-1}$$

If we restrict the coefficients we can guarrantee $n^{-\beta/2}$ for any $1 \leq \beta < 2$ provided we allow the coefficients to grow like $k^{2\beta k/(2-\beta)}$, namely, $A_k \geq k^{2\beta/(2-\beta)}$. Thus Jackson's estimate is obtained with $\beta = 1$ and we can get as close as we wish to n^{-1} by letting β approach 2. However it can be shown that (3) cannot be achieved for any β.

3 Monotone approximation with P_A

Let P_A^+ and P_A^1 denote the sets of nonnegative polynomials in P_A and the set of nondcreasing polynomials in P_A, respectively. Then the generalized Bernstein polynomials provide a constuctive proof for Weierstrass' type theorems on approximation with P_A^+ and P_A^1. To be precise we have

THEOREM 3. <u>The set P_A^+ is dense in the set of nonnegative functions in $C_o[0,1]$ and the set P_A^1 is dense in the set of monotone nondecreasing functions in $C_o[0,1]$ if and only if there exists a subsequence $\{k_i\}$ satisfying</u> (1).

<u>Proof</u>. That condition (1) is necessary follows by Roulier [6] exactly as it is necessary for the general case. As for the

sufficiency suppose (1) holds. Then for $f \in C_o[0,1]$ construct the generalized Bernstein polynomial of the sequence $\{k_i\}$ associated with f (see [5]). Given $\varepsilon > 0$ then for $n > N$

$$(4) \quad ||f - B_n(f,\cdot)|| < \varepsilon/2$$

where $B_n(f,\cdot)$ is the n-th generalized Bernstein polynomial. Now

$$(5) \quad B_n(f,x) = \sum_{j=1}^{n} f(\alpha_{nj})\, p_{nj}(x)$$

where

$$p_{nj}(x) = (-1)^{n-j}\, k_{j+1}\cdots k_n \sum_{i=j}^{n} x^{k_i}/\omega_{nj}'(k_i)$$

$$\omega_{nj}(x) = (x - k_j)\ \dots\ (x - k_n)$$

and

$$\alpha_{nj} = [(1 - k_1/k_{j+1})\ \dots\ (1 - k_1/k_n)]^{1/k_1} .$$

Since $f(0) = 0$ there exists a $\delta > 0$ such that $|f(\alpha_{nj})| < \varepsilon/2$ provided $\alpha_{nj} < \delta$. Thus if we put $m = m(n)$ to be such that $\alpha_{n,m-1} < \delta \leq \alpha_{nm}$, it follows by (4) and (5) together with the inequalities

$$p_{nj}(x) \geq 0, \qquad \sum_{j=1}^{n} p_{nj}(x) \leq 1$$

that

$$||f(x) - \sum_{j=m}^{n} f(\alpha_{nj})\, p_{nj}(x)|| < \varepsilon .$$

It has been shown in [3] that the polynomial

$$(6) \quad p_n(x) = \sum_{j=m}^{n} f(\alpha_{nj})\, p_{nj}(x)$$

belongs to P_A for all sufficiently large n. Now if f is

nonnegative in $[0,1]$, then so is p_n. Suppose then that f is nondecreasing. Then by [4] (2.2)

$$p'_{nj}(x) = x^{-1}[\lambda_j\, p_{nj}(x) - \lambda_{j+1}\, p_{n,j+1}(x)], \qquad 0 < x \leq 1.$$

Hence by (6) for $0 < x \leq 1$

$$\begin{aligned} p_n'(x) &= x^{-1}\sum_{j=m}^{n} f(\alpha_{nj})\,[\lambda_j\, p_{nj}(x) - \lambda_{j+1}\, p_{nj+1}(x)] \\ &= x^{-1}\sum_{j=m+1}^{n} \lambda_j\, p_{nj}[f(\alpha_{nj}) - f(\alpha_{n,j+1})] \\ &\quad + x^{-1}\lambda_m p_{nm}(x)\, f(\alpha_{nm}) \\ &\geq 0. \end{aligned}$$

The following result is obtained very easily from the fact that (1) is necessary and sufficient for P_A^+ to be dense in the nonnegative functions in $C_o[0,1]$. The reason we proved the second part of Theorem 3 is that it provides a better estimate on the rate of approximation (see Theorem 5).

THEOREM 4. *The set of polynomials p in P_A with $p^{(k)} \geq 0$ for some $k \geq 1$ is dense in the functions f in $C_o[0,1]$ with $f(0) = \ldots = f^{(k-1)}(0) = 0$, $f^{(k)} \geq 0$ if and only if (1) holds.*

We will conclude this note with an estimate on the rate of monotone approximation for some sequences A.

THEOREM 5. *Let A satisfy $A_k \geq \delta k$, $k \geq k_o$ for some $\delta > 0$. Then there exists a constant $K = K(\delta)$ independent of A such that for each nonnegative or nondecreasing $f \in C_o[0,1]$ and all sufficiently large n there is an n-th degree polynomial in P_A^+ or P_A^1, respectively, such that*

$$\|f - p_n\| \leq K\omega(f, n^{-1/2}).$$

References

1 Bak, J., M.v.Golitschek and D. Leviatan, The rate of approximation by means of polynomials with restricted coefficients, Israel J. Math. (to appear)

2 Golitschek, M.v., Permissible bounds on the coefficients of generalized polynomials, Proc. Conf. Approximation Theory, Austin, Texas 1973, G.G. Lorentz Editor, Academic Press, 1973.

3 Golitschek, M.v. and D. Leviatan, Permissible bounds on the coefficients of approximating polynomials with real or complex exponents, J. Math. Anal. and Appl. (to appear)

4 Leviatan, D., Tauberian constants for generalized Hausdorff transformations, J. London Math. Soc. 43 (1968), 308-314.

5 Lorentz, G.G., Bernstein Polynomials, University of Toronto Press, Toronto 1953.

6 Roulier, J.A., Restrictions on the coefficients of approximating polynomials, J. Approximation Theory 6 (1972), 276-282.

D. Leviatan
Department of Mathematics
Tel-Aviv University
Tel-Aviv, Israel.

REGULARITY OF SOME SPECIAL BIRKHOFF MATRICES

G. G. Lorentz, S. S. Stangler and K. L. Zeller

1 Introduction

We shall discuss the Birkhoff interpolation problem for algebraic polynomials of degree $\leq n$, for a set of knots X: $x_1 < x_2 < \ldots < x_m$ and for an incidence matrix $E = (e_{ik})_{i=1,\ k=0}^{m\quad n}$. Terminology of [6] and [3] will be used without explanation. Our main problem is to find when a given matrix E or a given pair E,X is regular (poised) or singular (nonpoised, or poisonous).

After an example in [4] of a three row matrix which disproved the "Atkinson-Sharma conjecture," much attention has been devoted to three row matrices E of the following type

$$(1.1)\quad e_{ik} = \begin{cases} 1 \text{ if } i = 1,\ 0 \leq k < p;\ i = 2,\ k = k_1,\ k_2; \\ \quad i = 3,\ 0 \leq k < q; \\ 0 \text{ otherwise} \end{cases}$$

In this case $n = p + q + 1$. Without loss of generality we shall assume that $p \leq q$, that E satisfies the Birkhoff condition and contains two supported separated ones: $0 < k_1$, $k_1 + 1 < k_2 < n$. If these inequalities are satisfied, we write $E = (p,q;k_1,k_2)$. Papers [1] and [2] describe independent investigations of DeVore, Meir and Sharma, and of Lorentz and Zeller about the regularity of these matrices. We shall "almost solve" the problem of regularity:

THEOREM 1. <u>A matrix</u> $E = (p,q;k_1,k_2)$ <u>can be regular only if</u>

$$(1.2)\quad k_1 + k_2 = p + q + 1, \qquad k_2 > q,$$

or if

(1.3) $p \le k_1 < k_2 \le q$;

in case (1.2) *it is regular if and only if* $p = q$; *in case* (1.3) *it may be regular and singular*, *but regularity of* $(p,q;k_1,k_2)$ *implies that of* $(p,q;k_1',k_2')$ if (a) $p \le k_1' \le k_1 < k_2 \le k_2' \le q$ *or if* (b) $k_1' = k_1 + 1$, $k_2' = k_2 + 1 \le q$.

Example in [4] is $E = (2,2;1,4)$, which is regular.

Our method, based on properties of zeros of polynomials, applies to many other cases. In particular, let E be an $(m - 1) \times (n + 1)$ Hermitian matrix with $n - 1$ ones

(1.4) $e_{ik} = 1,\ 0 \le k < r_i,\ i = 1,\dots,m-1,\ \Sigma r_i = n - 1$,

and let $X: x_1 < \dots < x_{m-1}$ be a fixed set of knots. In addition, let I be a "free" non-Hermitian row with ones in positions k_1, k_2, $0 < k_1$, $k_1 + 1 < k_2 < n$. We shall say that row i of E (or the knot x_i) is a *barrier* for I, if $r_i \ge k_2$.

Adding I to the rows of E in an arbitrary position, with the corresponding knot x, we obtain a pair E', X'. In this process, I is allowed to coincide with one of the rows i, provided i is not a barrier. Then two new ones are added to row i, in positions k_1, k_2 if $k_1 \ge r_i$ or r_i, k_2, if $k_1 < r_i$.

The following polynomials are essential:

(1.5) $$P(x,\lambda) = (x - \lambda) \prod_{i=1}^{m-1} (x - x_i)^{r_i} = (x - \lambda)Q(x);$$

here λ, $-\infty \le \lambda \le +\infty$ is a parameter. If $\lambda = \pm\infty$, P reduces

to Q. All zeros of a derivative $P^{(k)}$ are real. They are of two kinds: <u>fixed</u> zeros x_i, of multiplicity $r_i - k$, $i = 1,\ldots,m-1$, and the simple <u>variable</u> zeros, which depend on λ and can be obtained by Rolle's theorem. We shall denote the latter by

$$(1.6)\quad y_1^{(k)}(\lambda) < y_2^{(k)}(\lambda) < \cdots < y_z^{(k)}(\lambda),$$

their number by $z = z(k)$. For the variable zeros of $P^{(k_1)}$, $P^{(k_2)}$ we write simply $y_j'(\lambda)$, $y_j''(\lambda)$.

We say that the matrix E has property (R) if all pairs E', X' are regular.

THEOREM 2. (i) <u>A necessary condition for property</u> (R) <u>is that</u> E <u>should have at most one barrier</u>. (ii) <u>If</u> E <u>has exactly one barrier</u> i_o <u>then the necessary and sufficient condition for</u> (R) <u>is that</u> $z(k_1) = z(k_2) = z$ <u>and that for for each</u> $\lambda \neq x_{i_o}$, $\lambda \neq \infty$, <u>for some</u> s, $0 \leq s \leq z$

$$(1.7)\quad y_1'(\lambda) < y_1''(\lambda) < \cdots < y_s''(\lambda) < x_{i_o} < y_{s+1}''(\lambda) < \cdots < y_z'(\lambda).$$

(iii) <u>If</u> E <u>has no barriers, the necessary and sufficient condition is that</u> $z(k_1) = z(k_2) + 1 = z$ <u>and for each</u> $\lambda \neq \infty$,

$$(1.8)\quad y_1'(\lambda) < y_1''(\lambda) < \cdots < y_{z-1}''(\lambda) < y_z'(\lambda).$$

It should be noted that similar results have been given by Lorentz [2] in terms of zeros of $Q^{(k_1-1)}$, $Q^{(k_2-1)}$, rather than $P^{(k_1)}$, $P^{(k_2)}$. The difference with the present Theorem 2 is that conditions in [2] are only necessary; in contrast to (1.7) and (1.8) they do not contain λ; they apply to a slightly more general class of matrices; finally, if they are not

satisfied, apparently <u>strong</u> singularity of E follows. It would be desirable to combine the two different approaches.

2 <u>Some identities</u>

We shall use some identities valid for the polynomial

$$(2.1) \quad Q(x) = (x+1)^p (x-1)^q, \; 0 \le p \le q.$$

LEMMA 1. *For integers* $p + q = k + \ell$, $0 \le k \le \ell$, *the function* $Q(x)$ *satisfies the symmetry relation*

$$(2.2) \quad \frac{1}{\ell!} Q^{(\ell)}(-x) = (-1)^k (x+1)^{q-\ell} (x-1)^{p-\ell} \frac{1}{k!} Q^{(k)}(x),$$

$$x \ne \pm 1.$$

Proof. First assume that $k \le p \le q \le \ell$. Then by means of Leibniz' formula we obtain

$$\frac{1}{k!} Q^{(k)}(x) = \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} i! \binom{p}{i} (x+1)^{p-i} (k-i)! \binom{q}{k-1} (x-1)^{q-k+i}$$

$$(2.3) \quad = \sum_{i=0}^{k} \binom{p}{i} \binom{q}{k-i} (x+1)^{p-i} (x-1)^{q-k+i}$$

$$= (x+1)^{\ell-q} (x-1)^{\ell-p} \sum_{i=0}^{k} \binom{p}{i} \binom{q}{k-1} (x+1)^{k-i} (x-1)^{i},$$

because $p - k = \ell - q$, $q - k = \ell - p$. The similar sum for the derivative $Q^{(\ell)}$ has zero terms for $i < \ell - q$ and for $i > p$. Omitting them, we get

$$(2.4)\quad \frac{1}{\ell!}\, Q^{(\ell)}(x) = \sum_{i=\ell-q}^{p} \binom{p}{i}\binom{q}{\ell-i}(x+1)^{p-i}(x-1)^{q-\ell+i}.$$

Making substitution $i = p - j$, we derive

$$\frac{1}{\ell!}\, Q^{(\ell)}(x) = \sum_{j=0}^{k} \binom{p}{j}\binom{q}{k-j}(x+1)^{j}(x-1)^{k-j}$$

and therefore

$$(2.5)\quad \frac{1}{\ell!}\, Q^{(\ell)}(-x) = (-1)^{k} \sum_{j=0}^{k} \binom{p}{j}\binom{q}{k-j}(x-1)^{j}(x+1)^{k-j}.$$

Comparing this with (2.3), we obtain the required formula.

The case $p \leq k \leq \ell \leq q$ is very similar. Here, however, the sum in (2.3) is only for the range $0 \leq i \leq p$; the sum in (2.4) is for the same range, as well as the sum in (2.5) after the substitution $j = p - i$.

It follows from (2.2) that the sets of zeros of $Q^{(k)}$ and $Q^{(\ell)}$ in $(-1,+1)$ are symmetric to each other with respect to the origin.

COROLLARY 1. *If* $k + \ell = 2p$, *the derivatives of* $Q(x) = (x^2 - 1)^p$ *satisfy the relation*

$$(2.6)\quad \frac{1}{\ell!}\, Q^{(\ell)}(x) = (x^2-1)^{p-\ell}\, \frac{1}{k!}\, Q^{(k)}(x).$$

This follows from (2.2) if one observes that the derivative $Q^{(\ell)}$ is even (or odd) if the integer ℓ is even (or odd). (See also [3]).

3 *Properties of polynomials*

We shall need the following well-known form of Rolle's theorem, compare [4], [7].

THEOREM 3. *For each* n *there is a* δ, $0 < \delta < 1/2$ *with the following property*. *If* $\alpha < \beta$, $\beta - \alpha \geq d$, *are two*

<u>adjacent real zeros of a polynomial</u> P_n <u>of degree</u> $\leq n$, <u>then</u> P'_n <u>has a zero in</u> $(\alpha + \delta d, \beta - \delta d)$.

In particular, if all zeros of P_n are real, then the intervals $(\alpha, \alpha + \delta d)$, $(\beta - \delta d, \beta)$ contain no zeros of P'_n.

We shall consider polynomials $P_n(x,\lambda)$ which depend continuously on a parameter λ. If all zeros of P_n are real, we write them as a sequence

$$y_1(\lambda) < y_2(\lambda) < \dots < y_n(\lambda).$$

It is easy to prove that each $y_i(\lambda)$ depends continuously upon λ. The same is true for the zeros of $P_n^{(k)}(x,\lambda)$.

THEOREM 4. <u>Let</u> $P(x) = \prod_{j=1}^{m} (x - x_j)^{r_j}$, $n \geq 2$, m, $r_j \geq 1$, $x_1 < \dots < x_m$ <u>be a polynomial of degree</u> n <u>with real zeros</u>.
(i) <u>The zeros of the derivative</u> P' <u>depend in a monotone fashion upon the</u> x_j: <u>they increase if all</u> x_j <u>increase.</u>
(ii) <u>If the multiplicity</u> r <u>of one of the zeros</u> $x_j = a$ <u>increases, while the remaining multiplicities do not change, then the</u> x'_j <u>move away from</u> a.

<u>Proof</u>. (i) The derivative P' has two kinds of zeros: zeros x_j of multiplicity $r_j - 1$, and Rolle zeros, x'_j, $x_j < x'_j < x_{j+1}$, or zeros of

$$(3.1) \quad R(x) = \frac{P'(x)}{P(x)} = \sum_{j=1}^{m} \frac{r_j}{x - x_j}.$$

The function R is decreasing between its poles x_j. We can assume that only one of the zeros of R, say $a = x_j$, of multiplicity r, moves to the right, to the position $a' > a$. This increase of a will make the term $r/(x - a)$ of (3.1) larger both for $x < a$ and for $x > a'$, in particular around any of x'_i. Consequently x'_i will move to the right.

(ii) Similarly, enlargement of r will make the term $r/(x - a)$ of (3.1) larger if $x > a$, smaller, if $x < a$. Consequently, a zero x_i' will move right if $x_i' > a$, and move left if $x_i' < a$.

COROLLARY 2. *Let*

$$P_i(x) = (x + 1)^{p_i}(x - 1)^{q_i}, \; i = 1, 2, \; k \leq p_1 \leq p_2, \; q_1 \geq q_2.$$

The Rolle zeros x_j', x_j'' *of* $P_1^{(k)}, P_2^{(k)}$ *in* $(-1,+1)$ *satisfy the inequalities*

(3.2) $\quad x_j' \leq x_j'', \; j = 1, \ldots, k.$

Proof. There are k Rolle zeros of the derivatives, and they lie in $(-1,+1)$. We first consider P_1', P_2', with one zero x_1', x_1'' each. We can obtain P_2 from P_1 by increasing p_1 and by decreasing q_1. Each of these operations moves x_1' to the right. This proves our statement for the first derivatives. After this, we apply the same argument to Rolle zeros of

$$P_i' = (x + 1)^{p_i - 1}(x - x_1)(x - 1)^{q_i - 1}, \; i = 1, 2$$

and so on.

COROLLARY 3. *Let* $P(x) = (x + 1)^p(x - 1)^q$, $k \leq p \leq q$. *Then the* k *Rolle zeros* x_j *of* $P^{(k)}$ *in* $(-1,+1)$ *satisfy*

(3.3) $\quad x_j \leq -x_{k+1-j}, \quad j = 1, \ldots, k.$

Proof. In Corollary 2 we take $P_1 = P$, $P_2(x) = (x + 1)^q(x - 1)^p = \pm P(-x)$. The zeros of P_2 are $-x_k < -x_{k-1} < \ldots < -x_1$, and we obtain (3.3) from (3.2).

We need also a result reminiscent of Markov's theorem. For a similar result and proof see [5].

THEOREM 5. Let all zeros of the polynomials L, M be real: simple zeros x_i, $i = 1, \ldots, n$ and zero 0 of multiplicity r for L, simple zeros y_i, $i = 1, \ldots, n$ and zero 0 of multiplicity s, $0 \le s < r$ for M. If the zeros alternate in the sense that for some p, $0 \le p < n$,

$$(3.4) \quad x_1 \le y_1 \le \cdots \le y_p < 0 < y_{p+1} \le x_{p+1} \le \cdots \le x_n,$$

then the zeros ξ_i, η_i of the derivatives L', M' alternate strictly:

$$(3.5) \quad \xi_1 < \eta_1 < \ldots < \eta_p < 0 < \eta_{p+1} < \ldots < \xi_n.$$

Proof. It is sufficient to consider the inequalities (3.5) in $(-\infty, 0]$. Assume first that for some $j < p$, $y_j < x_{j+1}$. We show that for $x \in I = (y_j, x_{j+1})$ we have $\lambda(x) < \mu(x)$, where

$$\lambda(x) = \frac{L'}{L}(x) = \sum_{i=1}^{n} \frac{1}{x - x_i} + \frac{r}{x}, \quad \mu(x) = \frac{M'}{M}(x)$$
$$= \sum_{i=1}^{n} \frac{1}{x - y_i} + \frac{s}{x}.$$

Indeed, on this interval we have

$$x - x_i \ge x - y_i, \ i = 1, \ldots, p; \ x > x - y_{p+1};$$

$$x - x_i \ge x - y_{i+1}, \ i = p + 1, \ldots, n - 1.$$

Right and left sides of each of the inequalities have the same sign on I, hence $1/(x - x_i) \le 1/(x - y_i)$, $i = 1, \ldots, p$; $1/x < 1/(x - y_{p+1})$; $1/(x - x_i) \le 1/(1 - y_{i+1})$, $i = p+1, \ldots, n-1$,

moreover $1/x < 0$, $1/(x - x_n) < 0$. Adding, we obtain $\lambda(x) < \mu(x)$ on I. The same argument yields $\lambda(x) < \mu(x)$ for $x \in (y_p, 0)$.

Selecting $j \leq p$, we now prove that $\xi_j < \eta_j$. We can assume that $y_j < \xi_j$ (otherwise $\xi_j \leq y_j < \eta_j$), and that $\eta_j < x_{j+1}$ (otherwise $\xi_j < x_{j+1} \leq \eta_j$), hence that ξ_j, $\eta_j \in I = (y_j, x_{j+1})$; for $j = p$ we have ξ_p, $\eta_p \in I = (y_p, 0)$. Then ξ_j, η_j are the only zeros of μ, λ on I. Since $\mu(x)$ decreases on I, it is positive on (y_j, η_j), hence the zero ξ_j of λ satisfies $\xi_j < \eta_j$. The inequality $\eta_j < \xi_{j+1}$ $(j < p)$ follows from $\eta_j < y_j \leq x_{j+1} < \xi_{j+1}$.

4 Proof of theorem 2

We have to study the dependence on λ of the zeros of the derivatives of polynomials (1.5). The number $z = z(k)$ of variable zeros (1.6) of $P^{(k)}(x,\lambda)$ does not depend on the position of λ for $\lambda \neq \infty$. This number can be easily computed if the r_i are given; we have $z(k) > 0$ for $0 \leq k < n$. The zeros $y_j^{(k)}(\lambda)$ depend on λ continuously, and by Theorem 4(i), strictly increase with λ.

Let $r_i \geq k$, then from Rolle's theorem it follows that there is a zero $y_j^{(k)}$ in the interval (x_i, λ) for $\lambda > x_i$; if λ is close to x_i, there is only one such zero, by Theorem 3. As λ moves to the left across x_i, so does the zero, and λ overtakes the zero. If $\lambda < x_1$, there may be at most one zero $y_1^{(k)} < x_1$. If λ is close to $-\infty$, such a zero must exist and be also close to $-\infty$, by Theorem 3. Similarly for λ close to $+\infty$. Let $q_1 < \dots < q_{z-1}$ be the Rolle zeros of $Q^{(k)}$. If λ is close to $-\infty$, the zeros (1.6) approach the points $-\infty$, $q_1, \dots, q_{z-1}$, for λ close to $+\infty$ they approximate $q_1, \dots, q_{z-1}$, $+\infty$. Thus, as λ moves from $+\infty$ to $-\infty$ and then across $-\infty$ again to $+\infty$, the zeros (1.6) perform a cyclic permutation, $y_j^{(k)}$ being replaced by $y_{j-1}^{(k)}$, and $y_1^{(k)}$ by $y_z^{(k)}$.

If there are two derivatives $P^{(k_1)}$, $P^{(k_2)}$ we have to consider zeros y_j', y_j'' of both. In this case, for a barrier x_i, Rolle's theorem and Theorem 3 reveal that in each small interval (x_i,λ), $\lambda > x_i$, there is exactly one pair y_j', y_{j_1}'' with the property $x_i < y_{j_1}'' < y_j' < \lambda$; similarly for (λ, x_i). Also, for λ close to $-\infty$, there are zeros $-\infty < y_1' < y_1''$, both close to $-\infty$, and similarly for λ close to $+\infty$.

The $P(x,\lambda)$ of (1.5) represent precisely all nontrivial polynomials annihilated by E,X. Let $X' = X \cup \{x\}$ with an arbitrary knot x, and let E' be the corresponding matrix with $(n + 1)$ entries one. We have

$$P^{(k_1)}(x,\lambda) = P^{(k_2)}(x,\lambda) = 0, \tag{4.1}$$

precisely when P is annihilated by E',X'. This is obvious if x is not one of the knots x_i, also when $x = x_i$ with $k_1 \geq r_i$. If $k_1 < r_i$, $x = x_i$ and x_i is not a barrier, then as a zero $y = y_j^{(k)}(\lambda)$ approaches x_i, Rolle's theorem produces a zero of $P^{(r_i)}$ between x_i and y, and we must have

$$P^{(r_i)}(x,\lambda) = P^{(k_2)}(x,\lambda) = 0, \tag{4.2}$$

which is then equivalent to (4.1). This proves that <u>matrix</u> E <u>has property</u> (R) <u>exactly when</u> (4.1) <u>cannot happen for</u> $-\infty < x$, $\lambda < +\infty$.

In case (i) of Theorem 2, there are two barriers i_1, i_2 with $r_{i_1}, r_{i_2} \geq k_2$. If $\lambda \notin (x_{i_1}, x_{i_2})$, this interval contains exactly k_1 zeros y' and k_2 zeros y''; for λ in the interval, these numbers increase by one. We perform several translations of λ from a position λ_0 close to $+\infty$ to the left, and over $-\infty$ back to λ_0. At each step, a pair ordered

$y' < y'' < x_{i_2}$ enters the interval from the right, and another pair $x_{i_1} < y'' < y'$ leaves it at left. Since $k_2 - k_1 \geq 2$, there are two zeros y'' contained between two adjacent y'. After a certain number of steps we obtain a contradiction to this fact, if the zeros y', y'' never cross, that is, if (4.1) is impossible. This shows that (R) must be violated.

In cases (ii) and (iii) we also obtain a contradiction if the distribution (1.7) or (1.8) is violated, and (4.1) never happens. For this purpose we translate λ several times from $+\infty$ to $-\infty$ (or from $-\infty$ to $+\infty$) and watch the pair y', y'' close to $-\infty$ (or to $+\infty$). This completes the proof.

We note that if E contains a row i, $1 < i < m - 1$, for which $k_1 < r_i$, $r_i + 1 < k_2$, then E does not have the property (R). Indeed, the matrix E' is singular for $x = x_i$ by a result in [4].

If E has only two rows, then by theorems of Pólya and Atkinson-Sharma, (4.1) or (4.2) can happen only for $x_1 < x < x_2$. Thus we get:

COROLLARY 4. Conditions of Theorem 2 are necessary and sufficient for the regularity of the matrix $E = (p,q;k_1,k_2)$.

5 Proof of theorem 1

The three row matrix $E = (p,q;k_1,k_2)$ we treat by means of Corollary 4. We can assume $x_1 = -1$, $x_2 = +1$, and have to consider the polynomials

$$(5.1)\quad P(x,\lambda) = (x - \lambda)(x + 1)^p(x - 1)^q.$$

Rolle's theorem easily gives the number of variable (Rolle) zeros of $P^{(k)}$:

$$(5.2)\quad z(k) = \begin{cases} k+1 & \text{if } 0 \le k \le p \\ p+1 & \text{if } p \le k \le q \\ p+q+1-k & \text{if } q \le k \le p+q+1. \end{cases}$$

We have:

LEMMA 2. <u>The matrix</u> $E = (p,q;k_1,k_2)$ <u>is singular if</u> $k_1 + k_2 = p + q + 1$, $k_1 \le p$, $k_2 > q$ <u>and</u> $p < q$.

This answers a question of DeVore, Meir and Sharma [1].

<u>Proof</u>. Assuming $x_1 = -1$, $x_2 = 1$, we consider the zeros y', y'' of the derivatives $P^{(k_1)}$, $P^{(k_2)}$ of the polynomial (5.1), when $-\infty < \lambda < 1$. We have $z(k_1) = k_1 + 1$, $z(k_2) = k_1$. First let $\lambda = -1$. Then $y_1'(-1) = -1$. Here $P = (x+1)^{p+1}(x-1)^q$. Since $k_1 + k_2 = (p+1) + q$, Lemma 1 applies: the zeros of $P^{(k_1)}$, $P^{(k_2)}$ <u>inside</u> $(-1,+1)$ are symmetric about 0. In particular, $y_1''(-1) = -y_{k_1+1}'(-1)$. Since $p + 1 \le q$, we can use (3.3) from Corollary 3: $y_2'(-1) \le -y_{k_1+1}'(-1)$. Together this gives $y_2'(-1) \le y_1''(-1)$. On the other hand, considerations of the first part of section 4 show that for λ close to $-\infty$, $y_1''(\lambda) < -1 \le y_2'(\lambda)$. Hence for some λ, $y_1''(\lambda) = y_2'(\lambda)$, and E is singular.

We consider different cases of Theorem 2.

<u>Case (i)</u>: <u>two barriers</u>: $k_2 \le p$. Here E is singular.

<u>Case (iii)</u>: <u>no barriers</u>: $k_2 > q$. As a necessary condition of regularity we have $z(k_1) = z(k_2) + 1 = z$. It follows from (5.2) that this can happen only if both $z = k_1 + 1$ for $k_1 \le p$ and $z = p + q + 2 - k_2$ for $k_2 > q$, hence if $k_1 + k_2 = p + q + 1$. Lemma 2 gives singularity of E if $p < q$. On the other hand, both in [1] and [3] it is shown that E

is regular if $p = q$.

Case (ii): one barrier: $p < k_2 \leq q$. Since $z(k_2) = p + 1$ in our range, the necessary condition $z(k_1) = z(k_2)$ gives that $p \leq k_1 < q$ also, hence $p \leq k_1 < k_2 \leq q$. Here E can be regular and singular. In fact, DeVore, Meir and Sharma [1] gave simple necessary and sufficient conditions for the regularity of E when $p = 1$.

To prove (a) of Theorem 1, we note that by (5.2), for $p \leq k < \ell \leq q$, $z(k) = z(\ell) = p + 1$, and by Rolle's theorem,

$$(5.3) \quad y_j^{(k)}(\lambda) < y_j^{(\ell)}(\lambda),$$

for the zeros left of $+1$. Condition (1.7) reads here

$$(5.4) \quad \begin{aligned} &y_1'(\lambda) < y_1''(\lambda) < \ldots < y_p''(\lambda) < 1 < y_{p+1}''(\lambda) < y_{p+1}'(\lambda), \quad 1 < \lambda < +\infty \\ &y_1'(\lambda) < y_1''(\lambda) < \ldots < y_{p+1}''(\lambda) < 1, \ -\infty < \lambda < 1. \end{aligned}$$

This order can be disturbed only on the left of 1; this happens precisely when for some λ and $i \leq j$,

$$(5.5) \quad y_j''(\lambda) \leq y_i'(\lambda).$$

If this holds for the pair k_1', k_2', $k_1' \leq k_1 < k_2 \leq k_2'$, then by (5.3) also for the pair k_1, k_2. The inequality (5.5) is reversed for λ close to $-\infty$. Hence $(p,q;k_1,k_2)$ is singular whenever $(p,q;k_1',k_2')$ is.

Statement (b) of Theorem 1 follows from Theorem 5.

References

1. DeVore, R., A. Meir, and A. Sharma, Strongly and weakly non-poised H-B interpolation problems, Canadian J. Math., 25 (1973), 1040-1050.

2. Lorentz, G. G., The Birkhoff interpolation problem: New methods and results, in: Proceedings Conference Oberwolfach, Birkhäuser Verlag, Basel 1974, pp. 481-501.

3. Lorentz, G. G. Birkhoff interpolation problem, Report CNA-103 of the Center for Numerical Analysis, The University of Texas, Austin, 1975, 117 pp.

4. Lorentz, G. G. and K. L. Zeller, Birkhoff interpolation, SIAM J. Numer. Anal., 8 (1971), 43-48.

5. Lorentz, G. G. and K. L. Zeller, Monotone approximation by algebraic polynomials, Trans. Amer. Math. Soc. 149 (1970), 1-18.

6. Lorentz, G. G. and K. L. Zeller, Birkhoff interpolation problem: Coalescence of rows. Arch. Math., (Basel), 26, (1975), 189-192.

7. Riesz, M., Eine trigonometrische Interpolations Formel und einige Ungleichungen für Polynome, Jahresber. Deutsch. Math. Verein., 23 (1914), 354-368.

G. G. Lorentz*
Department of Mathematics
RLM 8-100
The University of Texas
Austin, Texas 78712

S. S. Stangler, K. L. Zeller
Mathematisches Institut der
Universität Tübingen,
7400 Tübingen, BRD.

*Supported, in part, by Grant MPS 75-09833 of the National Science Foundation.

MÜNTZ-SZÁSZ TYPE APPROXIMATION RESULTS AND THE PALEY-WIENER THEOREM

W. A. J. Luxemburg

1 The Müntz-Szász theorem

S. N. Bernstein, inspired by the Weierstrass approximation theorem, raised in his prize-winning essay [1] of 1912 the following question: Given an increasing sequence $\Lambda = \{\lambda_0 = 0 < \lambda_1 < \lambda_2 < \ldots\}$ of nonnegative real numbers what are the necessary and sufficient conditions in order that the span of the powers $t^{\lambda_n}(n \geq 0)$ is dense in $C[0,1]$? He showed that the condition: $\Sigma((1 + \log\lambda_n)/\lambda_n : \lambda_n > 0) = \infty$ is necessary and that the condition: "The sequence Λ is bounded" is sufficient. Bernstein conjectured, however, that the span of the powers $t^{\lambda_n}(n \geq 0)$ is dense in $C[0,1]$ if and only if $\Sigma(1/\lambda_n : n \geq 1) = \infty$.

In 1914 Ch. Müntz [12] verified the conjecture. He used the at that time newly developed method of the Gram determinants to obtain formulas for the distance in $L^2[0,1]$ of the function t^λ to the span of the functions $t^{\lambda_k}(0 \leq k \leq n)$. The Gram determinants which arise in this manner are of the form $\det(1/(a_i + a_j + 1)), i, j = 0,1,2,\ldots,n$, which are known as Cauchy determinants and can be evaluated explicitly. This led to the important result that the distance $d(\lambda)(\lambda > 0)$ in $L^2[0,1]$ of the function t^λ to the linear span of the powers $t^{\lambda_n}(n \geq 0)$ can be expressed by the formula

$$(1.1)\quad d(\lambda) = \frac{1}{\sqrt{2\lambda + 1}} \prod_{n=0}^{\infty} \left(1 - \frac{2\lambda + 1}{\lambda_n + \lambda + 1}\right).$$

Hence, it follows from the Weierstrass approximation theorem that the span of the powers $t^{\lambda_n}(n \geq 0)$ is dense in $C[0,1]$ if and only if $d(n) = 0$ for all $n = 1,2,\ldots.$ This is obviously the case if and only if $\Sigma(1/\lambda_n : \lambda_n > 0) = \infty$. Thus proving the conjecture for the L^2-metric. Using Fejér's theorem concerning the Cesàro summability of the Fourier series of continuous functions Müntz was able to show that the L^2-result implies the same result for $C[0,1]$, and so settled the Bernstein conjecture.

In 1916 O. Szász [15] extended the result to certain sequences of complex numbers and at the same time was able to simplify that part of Müntz' proof in which he showed that the L^2-result implies the $C[0,1]$-result by simply observing that

$$(1.2)\quad |t^n - \Sigma a_i t^{\lambda_i}| = (n+1)\ \left|\int_0^t (x^{n-1} - \Sigma a_i' x^{\lambda_i - 1})dx\right| \leq$$
$$(n+1)\left(\int_0^1 |x^{n-1} - \Sigma a_i' x^{\lambda_i - 1}|^2 dx\right)^{1/2}.$$

There is still another way of approaching the Müntz-Szász theorem. Observing that the linear span of the powers $t^{\lambda_n}(n \geq 0)$ is dense in, say, $L^2[0,1]$ if and only if $f \in L^2[0,1]$ and

$$\int_0^1 t^{\lambda_n} f(t)dt = 0$$

for all $n \geq 0$ implies $f \equiv 0$, we discover that the Müntz-Szász theorem is equivalent to the problem of characterizing the set of uniqueness of functions analytic in $Re(z) > 0$ of the form

$$\int_0^1 t^z f(t)dt.$$

Under the substitution $t = \exp(-x)$ the integral

$$\int_0^1 t^z f(t)dt$$

transforms into an integral of the form

$$(1.3) \quad F(z) = \int_0^\infty f(\exp(-x))\cdot\exp(-x/2)\cdot\exp(-x(z+\tfrac{1}{2}))dx,$$

with $g(x) = f(\exp(-x))\cdot\exp(-x/2) \in L^2(0,\infty)$ and $\mathrm{Re}(z) > -1/2$. We conclude that the question whether the powers $t^{\lambda_n} (n \geq 0)$ span $L^2[0,1]$ is equivalent to the question whether the set $\Lambda = \{\lambda_n : n \geq 0\}$ of real numbers is a set of uniqueness of analytic functions defined by Laplace integrals of the form

$$(1.4) \quad F(z) = \int_0^\infty f(t)e^{-zt}dt, \ f \in L^2(0,\infty) \text{ and } \mathrm{Re}(z) > 0.$$

The basic idea of this method goes back to Szász [15] and has been used by various authors such as Carleman [2], Szegö [16], Paley-Wiener [13], p. 32-34 and notably L. Schwartz in his thesis [14].

To answer the question about what sets Λ are sets of uniqueness of (1.4) Schwartz used the following theorem of Blaschke. Let D be a domain in the complex plane and let $\{z_n : n \geq 0\}$ be a sequence of points in D. Then there exists a nonzero bounded analytic function in D which vanishes on the points $z_n (n \geq 0)$ if and only if $\Sigma(G(a,z_n): z_n \neq a) < \infty$, where G is the Green's function of D and a is an arbitrary point of D.

For a domain of the type $\mathrm{Re}(z) \geq \varepsilon$, $\varepsilon > 0$ the Blaschke condition is equivalent to $\lim \lambda_n = 0$ and $\Sigma(1/\lambda_n : \lambda_n > \varepsilon) < \infty$. Since for every $\varepsilon > 0$ and analytic function of the type (1.4) is bounded in $\mathrm{Re}(z) \geq \varepsilon$ it follows immediately that the

condition $\Sigma(1/\lambda_n : n > 0) = \infty$ is sufficient for Λ to be a set of uniqueness. To prove by this method that the condition is also necessary one has only to observe that if $\Sigma(1/\lambda_n : n > 0) < \infty$, then the function

$$(1 + z)^{-2} \prod_{n \geq 0} ((\lambda_k - z)/(\lambda_k + z))$$

can be written in the form (1.4) with $f \neq 0$ and $f \in L^p(0,1)$ for all $p \geq 1$.

Because of the elementary nature of the Gram determinant method the Müntz-Szász theorem for the unit interval is treated in most textbooks on approximation theory. If instead of the unit interval we consider an interval of the form $[a,b]$, $a > 0$ the analysis of the Bernstein problem becomes considerably more complicated. Of course, the condition $\Sigma(1/\lambda_n) = \infty$ is still sufficient for the span of the powers $\{t^{\lambda_n}\}$ to be dense in $L^p[a,b] (p \geq 1)$ and $C[a,b]$. But the fact that the condition is also necessary in this case turns out to be far from trivial. The main reason for this is the fact that for an interval $[a,b]$, $a > 0$ the Gram determinants can no longer be expressed in simple terms. Furthermore, in this case the complex variable method described above leads to another uniqueness problem.

For special sequences of integers the first proof that the conditions is also necessary in this case was presented in 1943 in a paper by Clarkson and Erdös [3]. They did not use the two methods described above directly but rather reduced this case by an ingeneous indirect argument to the case of the unit interval. This argument was refined later in a paper by J. Korevaar [8] to include more general sequences. Also L. Schwartz in his thesis [14] gave a complete treatment of the problem of approximating certain functions by sums of expoentials.

But also his treatment of the problems for arbitrary intervals was indirect. This somewhat unsatisfactory state of affairs may perhaps explain why the Müntz-Szász theorem for arbitrary intervals has not yet found its way into the textbooks.

Recently, however, W. A. J. Luxemburg and J. Korevaar [10], have shown that a direct approach via the complex variable method is very well possible. We shall now discuss this approach briefly. For a similar but somewhat different method we refer the reader to W. Frost [6].

For the sake of simplicity we shall assume that $\Lambda = \{\lambda_n : n \geq 0\}$ is an increasing sequence of positive real numbers tending to infinity. The method, however, works equally well for certain sequences of complex numbers. We wish to prove the following theorem.

THEOREM 1. (J. Clarkson-P. Erdős, L. Schwartz, J. Korevaar). *In order that the powers* t^{λ_n} $(n \geq 0)$ *span* $L^p[a,b]$, $p \geq 1$ *or* $C[a,b]$, $a > 0$ *it is necessary and sufficient that* $\Sigma(1/\lambda_n) = \infty$.

Proof. Since the sufficiency of the condition is an immediate consequence from the classical Müntz-Szász theorem we will have only to show that the condition is necessary. To this end, we have to show that if $\Sigma(1/\lambda_n) < \infty$ the functions t^{λ_n} $(n \geq 0)$ do not span the spaces $L^p\{a,b\}$, $p \geq 1$ $C[a,b]$ $a > 0$. By duality, this means that for instance the functions t^{λ_n} $(n \geq 0)$ do not span, say, $L^2[a,b]$ if and only if there exists a nonzero function $f \in L^2[a,b]$ such that

$$\int_a^b t^{\lambda_n} f(t)dt = 0 \quad \text{for all } n.$$

This fact suggests that we consider the function

$$F(z) = \int_a^b t^z f(t)dt,$$

$f \in L^2[a,b]$ and z complex. Under the substitution $t = be^{-u}$ the integral transforms into

$$b^{z+1} \int_0^{\beta} f(be^{-u})e^{-u}e^{-zu}du,$$

where $\beta = \log(b/a) < \infty$, and $f(be^{-u})e^{-u} \in L^2[0,\beta]$. If we replace z by iz, then we recognize that we are dealing with the Fourier transforms of a function of compact support, which can be conveniently written in the form

$$(1.5)\quad G(z) = \int_0^{\beta} g(t) \exp(+izt)dt, \; g \in L^2[0,\beta] \text{ and } z \text{ complex.}$$

G is an entire function of exponential type $\leq \beta$, which is bounded on the real axis. In addition, it follows from the celebrated Paley-Wiener theorem [13] that

$$\int_{-\infty}^{+\infty} |G(x)|^2 dx < \infty .$$

Furthermore, $G(i\lambda_n) = 0$ for all $n = 0,1,2,\ldots$. This brief deduction shows that in order to prove that the condition of Theorem 1 is also necessary we need to construct an entire function of the form (1.5) which is not identically equal to zero and which vanishes at the points $i\lambda_n$ $(n = 0,1,2,\ldots)$, provided $\Sigma(1/\lambda_n) < \infty$. The latter hypothesis suggests to consider first the function

$$F(z) = \prod_{n\geq 0} \left(1 - \frac{z}{i\lambda_n}\right),$$

which G has to contain as a factor. Now F is of exponential type but of type zero and can not be written in the form (1.5). Furthermore for $z = x$ is real, $|G(x)| = \Pi(1 + x^2/\lambda_n^2)^{1/2}$ which tends to infinity as $|x| \to \infty$. It is not difficult to

show, however, that

$$\int_0^\infty (\log|G(x)|/x^2)dx = \frac{\pi}{2}\,\Sigma(1/\lambda_n) < \infty.$$

We can now use the following fundamental result. For the proof we refer to [10].

THEOREM 2. Let $\Lambda = \{\lambda_n : n \geq 0\}$ be an increasing sequence of positive numbers and let $\omega = \omega(x)$, $x \geq 0$ be a positive increasing function. Then for each $\tau > 0$ there exists an entire function F of exponential type τ satisfying the conditions:

(1.6) $F(i\lambda_n) = 0$ for all $n = 0,1,2,\ldots$

and $|F(x)| \leq \exp(-\omega(|x|)$, x real if and only if

$$\Sigma(1/\lambda_n) < \infty$$

and

$$\int_1^\infty \omega(x)/x^2 dx < \infty .$$

In fact, if the above conditions are satisfied, then we can construct a decreasing sequence $\{\varepsilon_n : n \geq 1\}$ of positive real numbers such that $\Sigma\varepsilon_n = \tau$ and

$$F(z) = \prod_{n\geq 0} \left(1 - \frac{z}{i\lambda_n}\right) \times \prod_{n\geq 1} \cos(\varepsilon_n z)$$

satisfies the required conditions.

Returning now to the situation at hand we observe that we can apply Theorem 2 to the sequence $\Lambda = \{\lambda_n\}$ and to the increasing function $\omega(x) = \log|G(x)| + \sqrt{|x|}$ and take for τ

the constant $\beta/2$. Then we obtain a function $F(z) = \Pi(1 - z/i\lambda_n) \prod_{n>1} \cos(\varepsilon_n z)$ which is entire and of exponential type $\beta/2$ and which on the real x-axis satisfies the condition $|F(x)| \leq \exp(-\sqrt{|x|})$. Hence,

$$\int_{-\infty}^{+\infty} |F(x)|^2 dx < \infty ,$$

and so, by the Paley-Wiener theorem there exists a nonzero function $f \in L^2[0,\beta]$ such that

$$F(z) \exp(i(\beta/2)z) = \int_0^\beta \exp(izt)f(t)dt$$

and which vanishes on the sequence Λ. Furthermore, since $|F(x)| \leq \exp(-\sqrt{|x|})$ it can even be shown that f is a C^∞-function, which settles the problem for all L^p-spaces ($p \geq 1$) and C. This completes the proof.

The direct approach via Theorem 2 of the necessity of the condition in Theorem 1 has various advantages. For instance it can be used to show directly that if $\Sigma(1/\lambda_n) < \infty$, then the distance of t^{λ_n} to the linear span of the powers $t^{\lambda_k} (k \neq n)$ is asymptotically $\geq (b - \varepsilon)^{\lambda_n}$, provided that the sequence $\{\lambda_k\}$ is well-spaced in the sense that $|\lambda_p - \lambda_q| \geq |p - q| \cdot d$, where $d > 0$. For this result and its consequences we have to refer to [10].

2 Extensions of the Müntz-Szász theorem

It is a natural question to ask. What is the form of the Müntz-Szász theorem if we replace the interval $[a,b]$ ($a \geq 0$) by a Jordan arc in the plane? The oldest result of this kind is due to J. L. Walsh [17] who showed that if γ is a Jordan arc, then the powers z^n ($n = 0,1,2,\ldots$) span the Banach space $C(\gamma)$ of all the continuous and complex functions on γ. Recently it was shown by J. Korevaar [9] and P. Malliavin and

J. A. Siddiqi [11] independently that if $\Lambda = \{\lambda_n\}$ is an increasing sequence of positive real numbers and if γ is an analytic arc, then $\Sigma(1/\lambda_n) < \infty$ implies that the exponentials $\exp(\lambda_n z)$, $z \in \gamma$ and $n = 0,1,2,\ldots$ fail to span $C(\gamma)$.

This result is obviously equivalent to the statement that if $\Sigma(1/\lambda_n) < \infty$, then there exists a nonzero measure μ on γ such that the entire function

$$F(z) = \int_\gamma \exp(zt)\,d\mu(t)$$

vanishes on Λ.

In [9] as well as in [11] it is shown that this theorem is a consequence of the fact that on an analytic arc certain functions can be constructed which belong to a quasi-analytic class on γ. It would be of interest to determine whether the complex variable method discussed in section 1 could be made to work in this case. This would amount to settling the following question. If $\Sigma(1/\lambda_n) < \infty$ does there exist a simple construction of an entire function G such that $F(z) = G(z)\Pi(1 - z/\lambda_n)$ can be written in the form

$$\int_\gamma \exp(zt)\,d\mu(t),$$

where μ is some nonzero measure on γ? A first step in settling this question would be to examine whether the Paley-Wiener theorem can be extended to include analytic arcs. At this time only the following partial answer to this problem is available.

Let F be an entire function of exponential type, let $h_F(\theta)(0 < \theta \leq 2\pi)$ denote its Phragmèn-Lindelöf function, let Φ denote its Borel transform and let S_Φ denote the conjugate indicator diagram of F.

The classical Paley-Wiener theorem states that if F belongs to a class L^2 on the real line, then S_Φ is a segment and Φ is of class E^2 on the domain determined by the complement of S_Φ (for the definition of the class E^2 we refer the reader to [5]). This formulation of the Paley-Wiener theorem suggests that the kind of generalization we are looking for is to determine necessary and sufficient conditions on F in order that Φ is of the class E^2 on the domain determined by the complement of the convex hull of a given analytic arc. In this direction we have the following result.

THEOREM 3. (V. E. Kacnel'son [7]), W. K. Delaney [4] and W. A. J. Luxemburg). *If F is an entire function of exponential type and if Φ belongs to the class E^2 on the complement of the convex hull Γ of a Jordan arc, then there exists a constant $K > 0$ which depends only on Γ such that*

$$(2.1)\quad \int_0^\infty |F(r\exp(i\theta))|^2 \exp(-2rh(\theta))\,dr \leq K \int_\gamma |\Phi(w)|^2 |dw|$$

for all $0 \leq \theta < 2\pi$, where h is the Minkowski support function of Γ and γ is the boundary of Γ. Furthermore, the converse does not hold, that is, there exists convex compact sets Γ such as disks, and entire functions F of exponential type for which the left-hand side of (2.1) is a finite and continuous function of θ but for which the right-hand side fails to be finite.

A detailed discussion of this result and its relation to the Müntz-Szász theorem will appear elsewhere.

References

1 Bernstein, S. N. Sur l'ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné, Classe des sciences de l'Académic Royale de Belgique, vol. IV (II) (1912), 104.

2 Clarleman, T., Über die Approximation analytischer Funktionen durch lineare Aggregate von vorgegebene Potenzen, Ark. Mat. Astr. Fys. 17 (1922), no. 9

3 Clarkson, J. A. and P. Erdös, Approximation by polynomials, Duke Math. J., 10 (1943), 5-11.

4 Delaney, W. K., A counterexample in the theory of Fourier transforms in the complex domain, Thesis (1975), California Institute of Technology.

5 Duren, P. L., Theory of H^p-Spaces, Academic Press, New York and London 1970.

6 Frost, W., Ein funktionentheoretischer Beweis des Satzes von Müntz, Manuscripta Math. 3 (1970), 357-374.

7 Kacnel'son, V. E., Generalization of the Paley-Wiener theorem of entire functions of finite degree, Teor. Funk. Funkcional Anal. i Prilozen 1 (1964), 99-110.

8 Korevaar, J., A characterization of the submanifold of C[a,b] spanned by the sequence $\{x^{n_k}\}$, Kon. Nederl. Ak. Wetensh. Proc. Ser. A 50 (1947), 750-758 = Ind. Math. 9 (1947), 360-368.

9 Korevaar, J., Approximation on curves by linear combinations of exponentials, Approximation theory (editor G. G. Lorentz). Acad. Press, New-York and London 1973, 387-393.

10 Luxemburg, W. A. J. and J. Korevaar, Entire functions and Müntz-Szász type approximation. Trans. Amer. Math. Soc., 157 (1971), 23-37.

11 Malliavin, P. and J. A. Siddiqi, Approximation polynomiale sur un arc analytique dans le plans complexe, C. R. Acad. Sci. Paris, 273 (1971), 105-108.

12 Müntz, C. H., Über den Approximationssatz von Weierstrass, H. A. Schwarz Festschrift, Berlin 1914, 303-312.

13 Paley, R. E. A. C. and N. Wiener, Fourier transforms in the complex domain, Amer. Math. Soc. Colloq. Publ., 19 (1934).

14 Schwartz, L., Étude des sommes d'exponentielles réelles, Act. Sc. Ind. 959, Hermann, Paris, 1943.

15 Szász, O., Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen, Math. Ann. 77 (1916), 482-496.

16 Szegö, G., Über dichte Funktionenfamilien, Ber.Math. Phys. Kl. der sächsischen Ak. der Wiss., 78 (1926), 373-380.

17 Walsh, J. L., Über die Entwicklung einer Funktion einer komplexen Veränderlichen nach Polynomen, Math. Ann. 96 (1927), 437-450.

W. A. J. Luxemburg*
California Institute of Technology
Department of Mathematics
Pasadena, California 91125

*Supported in part by Grant MS74-17845 of the National Science Foundation.

INTERPOLATION TO BOUNDARY DATA IN TRIANGLES WITH APPLICATION TO COMPATIBLE FINITE ELEMENTS

Lois Mansfield

Error bounds in Sobolev norms are given for the interpolation scheme presented in [1] and [2]. Applications of this scheme to finite element methods are given.

In [1] and [2] an interpolation scheme which interpolates to compatible boundary data consisting of values and normal derivatives of order up to m-1 given on the edges of a triangle T is introduced and analyzed. Suppose T has vertices $P_i = (x_i, y_i)$, i = 1,2,3. Let E_i be the edge opposite the vertex P_i. The interpolant $Q^m u$ of [1] and [2] may be expressed as

$$(1)\quad Q^m u = \frac{1}{2}[\mathcal{P}_1^m + \mathcal{P}_2^m + \mathcal{P}_3^m - \mathcal{L}^m]u,$$

where $\mathcal{L}^m = \mathcal{P}_i^m \mathcal{P}_j^m \mathcal{P}_k^m$, $i \neq j$, $j \neq k$, $k \neq i$, the product being taken in any order, and the projectors $\mathcal{P}_i^m$ are the Hermite interpolants to values and the first m-1 directional derivatives (parallel to E_i) on E_j and E_k interpolated along parallels to E_i. These directional derivatives can be uniquely computed from the given values f_j and normal derivatives $g_{j,i}$, j = 1,2,3; i = 1,...,m-1.

By Lemma 4.2 of [2], $Q^m u$ interpolates given values f_j and normal derivatives $g_{j,i}$ if and only if this data satisfies the compatibility conditions

$$(2)\quad \frac{\partial^k}{\partial s_i^k}\left(\frac{\partial^\ell u}{\partial s_j^\ell}\right)(P_\mu) = \frac{\partial^\ell}{\partial s_j^\ell}\left(\frac{\partial^k u}{\partial s_i^k}\right)(P_\mu), \qquad 0 \le k,\ell \le m-1,$$

at each vertex P_μ with adjacent sides E_i and E_j, where $\partial/\partial s_i$ denotes directional differentiation parallel to the i^{th} side.

Let T_{2m-1} be the $3m^2$-dimensional set of polynomials which are of degree 2m-1 along parallels to the three sides of T. By Lemma 4.1 of [2], $L^m u$ is the element of T_{2m-1} which interpolates u with respect to the interpolation conditions of (2).

The form of $Q^m u$ is a sum of directional derivatives of u evaluated on ∂T, each multiplied by a rational function. These rational functions are often referred to as "blending functions", and hence the interpolation scheme just described is often called blended interpolation.

In [2], it is shown that $Q^m q = q$ for all polynomials of degree 3m-1 or less. In addition, error estimates are given in the sup norm for the approximation of $u \in C^r$, $2m \leq r \leq 3m$, by its blended interpolant $Q^m u$, using the Sard kernel theorem. No estimates, however, were given concerning the approximation of derivatives of u by corresponding derivatives of Q^m. Such estimates are obtained in the following theorem.

THEOREM 1. <u>Let $Q^m u$ be the blended interpolant to $u \in H^r(T)$,</u> $2m \leq r \leq 3m$, <u>given by (1).</u> <u>Then</u>

$$(3) \quad ||u - Qu||_{H^k(T)} \leq \frac{Kh^r}{\rho^k} ||u||_{H^r(T)}, \qquad 0 \leq k \leq m,$$

<u>where h is the length of the longest side of the triangle T, ρ is the diameter of the inscribed circle, and K is a constant independent of u, h, and ρ.</u>

SKETCH OF PROOF. Using standard techniques in finite element analysis, we map T onto the standard triangle T_s with vertices (0,0), (1,0), and (0,1) using an affine transformation. This gives

$$(4) \quad ||u - Q^m u||_{H^k(T)} \leq \frac{C|J|^{1/2}}{\rho^k} ||\hat{u} - \hat{Q}^m \hat{u}||_{H^k(T_s)}$$

where J is the Jacobian of the transformation. Let $\{A_i\}_{i=1}^N$, $N = \binom{r}{2}$, form a unisolvent set such that $A_i \in T_s$ for each i.

The norm

$$\|\hat{v}\|^2_{r,2,T_s} = |\hat{v}|^2_{r,2,T_s} + \sum_{i=1}^{N} |\hat{v}(A_i)|^2$$

is known to be equivalent to the usual Sobolev norm. Let $\hat{q}$ be the unique polynomial of degree r-1 which interpolates $\hat{u}$ with respect to the points A_i. Then

$$\hat{u} - \hat{Q}^m\hat{u} = \hat{u} - \hat{q} - \hat{Q}^m(\hat{u} - \hat{q}).$$

We then show that $\hat{Q}^m \in L(H^r(T_s); H^m(T_s))$. It is clear that this is the case for $\hat{L}^m$. For $\hat{P}^m_i$, it is necessary to expand various derivatives of $\hat{u} - \hat{q}$ in Taylor series along the edges of T_s. Thus we have

$$(5)\quad ||\hat{u} - \hat{q} - \hat{Q}^m(\hat{u} - \hat{q})||_{H^k(T_s)} \leq C||\hat{u} - \hat{q}||_{H^r(T_s)}$$

$$\leq C'\|\hat{u} - \hat{q}\|_{r,2,T_s} = C'|\hat{u}|_{r,2,T_s} \leq h^r|J|^{-1/2}||u||_{H^r(T)},$$

which when combined with (4) proves the theorem.

In proving that $\hat{Q}^m \in L(H^r(T_s); H^m(T_s))$, one obtains the following useful inequality.

LEMMA 1. <u>Let $\hat{u} \in H^r(T_s)$</u>, $2m \leq r \leq 3m$. <u>Then</u>

$$(6)\quad ||\hat{Q}^m\hat{u}||_{H^k(T_s)} \leq C \max_{\substack{0\leq|\beta|\leq m-1 \\ 0\leq j\leq L}} \left(\int_{\partial T_s} \left[\frac{d^j}{d\sigma^j}(D^\beta\hat{Q}^m\hat{u})\right]^2 d\sigma\right)^{1/2},$$

$0 \leq k \leq m$, <u>where $L = \max\{m, m+2-|\beta|\}$.</u>

Thus $||\hat{Q}^m\hat{u}||_{H^m(T)}$ can be bounded in terms of derivatives of the given boundary data. We shall give an application of this inequality below.

One may obtain finite elements by taking the boundary data to be polynomials. The following result is of interest.

THEOREM 2. (Theorem 3 of [3]) <u>The interpolant $Q^m u$ defined above is a polynomial of degree k or less if and only if the</u>

boundary values f_j and normal derivatives $g_{j,i}$, $i = 1,\ldots,m-1$; $j = 1,2,3$, are polynomials of degrees k and $k-i$ respectively which satisfy the compatibility conditions (2).

The compatibility conditions (2) can be used to show that polynomial C^1-finite elements must be of degree at least five. Let τ be a triangulation of a polygonal region Ω of the plane such that two triangles are either disjoint or have a common vertex or common edge. A polynomial C^1-finite element subspace $S(\tau,k,1)$ is a subspace of piecewise polynomials which are globally in C^1 with values defined locally on each $T \in \tau$. In particular, $S(\tau,k,1)$ satisfies the properties: (i) a change in the value of any $q \in S(\tau,k,1)$ at a vertex P_i affects the value of q only on triangles having P_i as a common vertex, and (ii) on any edge the values and normal derivatives of $q \in S(\tau,k,1)$ are independent.

LEMMA 2. *Let $S(\tau,k,1)$ satisfy properties (i) and (ii). Then $k \geq 5$.*

PROOF. Let $T \in \tau$ have vertices P_0,P_1,P_2 labeled so that the angles at P_0 and P_1 are not right angles. Let $q \in S(\tau,k,1)$ have values f_j and normal derivatives g_j on the edge of T opposite P_j, $j = 0,1,2$. Since q is analytic in T, (2) must be satisfied. Thus

$$(7)\quad \begin{aligned} \alpha_1 g_1'(P_0) + \beta_1 f''(P_0) &= [\partial/\partial s_1(\alpha_1 \partial/\partial n + \beta_1 \partial/\partial s_1)q](P_0) \\ &= \partial/\partial s_1(\partial q/\partial s_2)(P_0) = \partial/\partial s_2(\partial q/\partial s_1)(P_0) \\ &= [\partial/\partial s_2(\alpha_2 \partial/\partial n + \beta_2 \partial/\partial s_2)q](P_0) \\ &= \alpha_2 g_2'(P_0) + \beta_2 f_2''(P_0). \end{aligned}$$

By properties (i) and (ii), a perturbation in the value of f_2 at P_1 must leave f_1, g_1, and g_2 invariant, and hence also $f_2''(P_0)$. A slightly simpler argument along these lines shows that $f_2(P_0)$ and $f_2'(P_0)$ must also remain invariant for any

perturbation in the value of $f_2(P_1)$. But this is only possible if f_2 can be defined to have prescribed values and prescribed first and second derivatives at P_0, and, by symmetry, at P_1. Thus f_2 must be of degree at least five.

This argument can easily be extended to $S(\tau,k,\ell)$, $\ell > 1$. The result contained in Lemma 2 had been obtained earlier by Ženíšek in [7]. Our proof is easy to generalize to higher dimensions. See [4].

The interpolation scheme defined by Q^m can be used to match exactly non-homogeneous boundary conditions given on the boundary of a polygon in the finite element approximation of solutions of elliptic differential equations. This has been considered by Mitchell and Marshall in their papers [5] and [6] and in the appendix of [3]. To obtain error bounds, one is led to consider the approximation of a function $u \in H^r(T)$, $2m \leq r \leq 3m$, by a blended interpolant $\bar{Q}^m u$ which interpolates to values (and possibly normal derivatives) on one side, say E_1, of a triangle T, and to polynomial approximations to u (and possibly normal derivatives) on the other two sides. Let us also assume that the triangulation τ is regular in the sense that for any $T \in \tau$, its smallest angle α satisfies $0 \leq \alpha_0 < \alpha$.

THEOREM 3. Suppose that on interior triangles finite elements are used for which one has the following error bounds:

$$(8) \quad \|u - \pi u\|_{H^m(T)} \leq K h^{r-m} \|u\|_{H^r(T)},$$

$$(9) \quad |u - \pi u|_{k,2,\partial T} \leq K h^{r-1-k} \|u\|_{H^{r-1}(\partial T)}, \quad \begin{matrix} 2m \leq r \leq 3m, \\ 0 \leq k \leq m, \end{matrix}$$

where πu is the interpolant to u from the space of finite elements with respect to some interpolation scheme. Then

$$(10) \quad \|u - \bar{Q}^m u\|_{H^m(T)} \leq K h^{r-m} \|u\|_{H^r(T)}.$$

SKETCH OF PROOF. We add and subtract πu and use (6). Partial derivatives of $\hat{\hat{Q}}^m(\hat{\pi}u - \hat{u})$ on the edge $\hat{E}_j$ can be expressed in terms of tangential derivatives and $\partial/\partial\nu_j$, the transform in T_s of $\partial/\partial n_j$ under the affine transformation which maps T onto T_s. Note that $\partial/\partial\nu_j$ depends upon the geometry of T, but this dependence can be expressed in terms of $(\sin\alpha)^{-1}$, which we have assumed is uniformly bounded away from zero for all $T \in \tau$. Thus

$$||\hat{\hat{Q}}^m(\hat{\pi}u - \hat{u})||_{H^m(T_s)}$$

$$\leq C \max_{\substack{0\leq i\leq m-1\\ 0\leq \ell\leq m-1\\ 0\leq j\leq L}} \left(\int_{\hat{E}_1} (\hat{g}_{1,i}^{(j+\ell-i)} - \hat{s}_{1,i}^{(j+\ell-i)})^2 d\sigma\right)^{1/2}$$

$$\leq C'|\hat{u}|_{r,2,T_s} .$$

References

1. Barnhill, R.E., G. Birkhoff and W.J. Gordon, "Smooth interpolation in triangles", J. Approximation Theory, 8 (1973), 114-128.

2. Barnhill, R.E., and L. Mansfield, "Error bounds for smooth interpolation in triangles", J. Approximation Theory, 11 (1974), 306-318.

3. Birkhoff, G. and L. Mansfield, "Compatible triangular finite elements", J. Math. Anal. Appl., 47 (1974), 531-553.

4. Mansfield, L., "Interpolation to boundary data in tetrahedra with applications to compatible finite elements", J. Math. Anal. Appl. (to appear).

5. Marshall, J.A., and A.R. Mitchell, "An exact boundary technique for improved accuracy in the finite element method", J. Inst. Math. Appl., 12 (1973), 355-362.

6. Mitchell, A.R., and J.A. Marshall, "Matching of essential boundary conditions in the finite element method", Proceedings of Numerical Analysis Conference at Dublin, 1974.

7. Ženíšek, A., "A general theorem on triangular finite

$C^{(m)}$-elements", R.A.I.R.O., 8 (1974), 119-127.

Lois Mansfield
Department of Computer Science
University of Kansas
Lawrence, Kansas 66045

CONVERGENCE RESULTS FOR CARDINAL HERMITE SPLINES

M.J. Marsden and S.D. Riemenschneider

By first investigating the behaviour and exact limit properties of the Fourier transforms of the fundamental splines for cardinal Hermite interpolation, results concerning the convergence as the degree tends to infinity can be readily obtained.

Let $S_{2m,r}$ denote the class of functions $S(x)$ such that $S(x) \in C^{2m-1-r}(R)$, and $S(x) \in \pi_{2m-1}$ in each interval $[\nu,\nu+1]$, $\nu \in Z$ (Z denotes the integers). Given an r-tuple of sequences, $(y^o,\cdots,y^{r-1})$, $y^s = \{y_\nu^s : \nu \in Z\}$, the cardinal Hermite interpolation problem is to find an (unique) element $S(x) \in S_{2m,r}$ such that $S^{(s)}(\nu) = y_\nu^s$ for $s = 0,1,\cdots,r-1$ and $\nu \in Z$. P. Lipow and I.J. Schoenberg [3] established that if the sequences are of power growth, i.e. $y_\nu = O(|\nu|^\gamma)$ as $|\nu| \to +\infty$, then there is a unique spline $S(x) \in S_{2m,r}$ that solves the interpolation problem and is of power growth, $|S(x)| = O(|x|^\gamma)$ as $|x| \to \infty$. Moreover, this spline is given by the formula

$$(1)\quad L_{2m,r}(y^o,\cdots,y^{r-1})(x) = \sum_{s=o}^{r-1} \sum_{\nu\in Z} y_\nu^s\, L_{2m,r,s}(x-\nu)$$

where the fundamental splines $L_{2m,r,s}(x)$ are uniquely determined by $L_{2m,r,s}^{(\rho)}(0) = \delta_{\rho,s}$, and $L_{2m,r,s}^{(\rho)}(\nu) = 0$, $\nu \neq 0$, $\rho = 0,1,\cdots,r-1$.

Recently, S.L. Lee [2] has established the Fourier transform representation of the fundamental splines as

Preceding page blank

$$(2) \quad L_{2m,r,s}(x) = \frac{(-i)^s}{2\pi} \int_{-\infty}^{+\infty} H_{2m,r,s}(u) e^{iux} du$$

where $H_{2m,r,s}(u)$ is the quotient of two Hankel determinants. More precisely, the denominator is the determinant of a matrix whose (i,j)-entry is $\Sigma_{k\in Z}[u + 2\pi k]^{-2m+(i-1)+(j-1)}$, $i,j = 1,\cdots,r$, while the numerator of $H_{2m,r,s}$ is the determinant of the matrix obtained from the denominator matrix by replacing the $(s+1)^{st}$-column by the column vector $(u^{-2m},u^{-2m+1},\cdots,u^{-2m+r-1})^T$.

The first result is a technical lemma that makes the dominant terms of $H_{2m,r,s}$ more visible.

LEMMA 1. *If* u/π *is not an integer, then*

$$(3) \quad H_{2m,r,s}(u) = \frac{\displaystyle\sum_{(i)} \frac{V(i)\ V_1(i)\ V_2(i)}{\prod_{1\le j\le s+1} (x+i_j)^{2m} \prod_{s+2\le j\le r} (x+i_j)^{2m-s-1}}}{\displaystyle(2\pi)^s \sum_k V^2(k) \prod_{1\le j\le r} (x+k_j)^{-2m}}$$

where $x = (u-2\pi\ell)/2\pi$ *and* ℓ *satisfies* (a) $2\ell+r$ *is an odd integer and* (b) $|u-2\pi\ell|<\pi$; (i) *denotes a distinct set* $i_1,\cdots,i_r$ *satisfying* (c) $2i_j+r$ *is an odd integer*, (d) $i_1 < i_2 < \cdots < i_s$, (e) $i_{s+1} = \ell$, *and* (f) $i_{s+1} < \cdots < i_r$; *the denominator sum is over distinct* $k_1 < k_2 < \cdots < k_r$ *satisfying* $2k_j+r$ *is an odd integer; and* $V(i) = V(i_1,\cdots,i_r)$, $V_1(i) = V(i_1,\cdots,i_s)$, $V_2(i) = (i_{s+2},\cdots,i_r)$ *and* $V(k) = V(k_1,\cdots,k_r)$ *are Vandermonde determinants*.

With the aid of the representation (3), the following properties of $H_{2m,r,s}(u)$ can be deduced by a "pairing" argument on dominant terms.

LEMMA 2. (a) $H_{2m,r,s}(u) = \mathcal{O}\left(\min(1,(r\pi/u)^{2m})\right)$ *as* $m \to \infty$.

(b) $\mathrm{Var}\{u^\rho H_{2m,r,s}(u)\} = \mathcal{O}(1)$ as $m \to \infty$ where $\rho = 0,1,2,\cdots,$ and the constants depend only on ρ, r and s.

Since $|x| \leq 1/2$ in (3), the dominant term of the denominator is given by the set $\eta = (\eta_1,\cdots,\eta_r)$ where $\eta_j = j-(r+1)/2$. In pairing this term with the numerator terms, it can be shown that in the limit as $m \to +\infty$ the only remaining terms are those for which (i) coincides with η as sets. Defining

$$(4)\quad Q_{r,s}(u) = (2\pi)^{-s} \sum_{(i)=\eta} V(i)V_1(i)V_2(i)V^{-2}(\eta) \prod_{s+2\leq j\leq r} (x+i_j)^{s+1}$$

on the interval $|u-2\pi\ell|<\pi$ where $|\ell|\leq(r-1)/2$ and $x = (u-2\pi\ell)/2\pi$, $Q_{r,s}(u) = 0$ for $|u|>r\pi$, and $Q_{r,s}(2\pi\eta_j+\pi)$ $= [Q_{r,s}(2\pi\eta_j+\pi-0)+Q_{r,s}(2\pi\eta_{j+1}-\pi+0)]/2$, $j=0,1,\cdots,r$, we have

PROPOSITION 3. The relation

$$(5)\quad \lim_{m\to\infty} H_{2m,r,s}(u) = Q_{r,s}(u)$$

holds for all u. Moreover, the convergence is uniform outside any neighbourhood of the points $2\pi\eta_j+\pi$, $j = 0,1,\cdots,r$.

Letting F denote the Fourier transform, Lemma 2(a), Proposition 3 and (2) combine to yield

LEMMA 4. The relation

$$(6)\quad \lim_{m\to\infty} L^{(\rho)}_{2m,r,s}(x) = (-i)^s\, F(Q_{r,s})^{(\rho)}(x)$$

holds uniformly in x for each integer $\rho \geq 0$.

If $\ell^p\oplus\cdots\oplus_r \ell^p$ denotes the Banach space of r-tuples $(y^o,\cdots,y^{r-1})$ of ℓ^p sequences with norm $\Sigma\|y^i\|_{\ell^p}$, and $W^{p,r-1}(R)$ denotes the Sobolev space of functions f for which $f^{(r-2)}$ is absolutely continuous and $f^{(j)} \in L^p(R)$,

$j = 0,1,\cdots,r-1$ with the norm $\Sigma \|f^{(j)}\|_{L^p}$, then Lipow and Schoenberg [3] have shown that $L_{2m,r}$ is a linear operator from $\ell^p \oplus \cdots \oplus_r \ell^p$ to $W^{p,r-1}(R)$. Following [4], Lemma 2(b) and Proposition 3 can be used to obtain convergence results in these spaces.

THEOREM 5. The operator $L_{2m,r}$ defined by (1) is a bounded linear operator from $\ell^p \oplus \cdots \oplus_r \ell^p$ to $W^{p,r-1}(R)$, $1 < p < +\infty$, whose norm is bounded independently of m.

THEOREM 6. If $W_r(y^o,\cdots,y^{r-1})(x) = \Sigma_{s=o}^{r-1} \Sigma_{\nu=-\infty}^{+\infty} y_\nu^s (-i)^s F(Q_{r,s})(x-\nu)$, then the relation

$$(7) \quad \lim_{m \to \infty} \| L_{2m,r}(y^o,\cdots,y^{r-1}) - W_r(y^o,\cdots,y^{r-1}) \|_{p,r-1} = 0$$

holds for each $(y^o,\cdots,y^{r-1}) \in \ell^p \oplus \cdots \oplus_r \ell^p$, $1 < p < \infty$.

If the data $\{y^s\}$ is given by $f^{(s)}(\nu) = y_\nu^s$, $\nu \in Z$, $s = 0,\cdots,r-1$, where f is a "nice" function, then uniform convergence results can be expected in light of Schoenberg [5].

THEOREM 7. Let $f(x)$ be defined by

$$(8) \quad f(x) = \sum_{s=o}^{r-1} \frac{(-i)^s}{2\pi} \int_{-r\pi}^{r\pi} Q_{r,s}(u) e^{ixu} d\tilde{\beta}_s(u)$$

where $\tilde{\beta}_s$ is the 2π-periodic extension of a bounded measure on $[-\pi,\pi)$. The relation

$$(9) \quad \lim_{m \to \infty} L_{2m,r}(f,f',\cdots,f^{(r-1)})^{(\rho)}(x) = f^{(\rho)}(x)$$

holds uniformly in x for each integer $\rho \geq 0$.

The proof follows from an identity which allows the identification $f^{(j)}(\nu) = f_j(\nu)$, $\nu \in Z$ where $f_j(x) = 1/2\pi \int_{-\pi}^{\pi} e^{ixu} d\beta_j(u)$. Then a summability argument (see [1] and [6]) yields

$$L_{2m,r}(f,f',\cdots,f^{(r-1)})^{(\rho)}(x)$$

$$= \sum_{s=o}^{r-1} \frac{(-i)^s}{2\pi} \int_{-\infty}^{\infty} H_{2m,r,s}(u)(iu)^{\rho} e^{ixu} d\tilde{\beta}_s$$

and the convergence properties of $H_{2m,r,s}$ provides the result.

Using the same sort of argument provides the theorem

THEOREM 8. The following classes of functions are equivalent for $1 < p \leq 2$:

(a) $W^{p,r-1}(\mathbb{R}) \cap \{f(x):\ f$ is of the form (8)$\}$

(b) $\{f(x):\ f(x) = W^{p,r-1}\text{-}\lim L_{2m,r}(y^o,\cdots,y^{r-1})(x),$
$(y^o,\cdots,y^{r-1}) \in \underbrace{\ell^p \oplus \cdots \oplus \ell^p}_{r}\}$

(c) $\{f(x):\ f(x) = W_r(y^o,\cdots,y^{r-1})(x),\ (y^o,\cdots,y^{r-1}) \in \underbrace{\ell^p \oplus \cdots \oplus \ell^p}_{r}\}$.

In conclusion, some explicit expressions for $Q_{r,s}(u)$ are noted: (r=2)

$$Q_{2,0}(u) = (2\pi-|u|)/2\pi,\ |u| \leq 2\pi \text{ and } Q_{2,1}(u) = \operatorname{sgn} u,$$

for $0 < |u| < 2\pi$;

(r=3) $Q_{3,0}(u) = (2\pi)^{-2}M_3(u) - (1/8)M_3''(u)$

$$Q_{3,1}(u) = -(2\pi)^{-2}M_3'(u) \text{ and } Q_{3,2}(u) = -(2\pi)^{-2}M_3''(u)/2$$

for $0 \leq |u| \leq 3\pi$ except $u = \pm\pi,\ \pm 3\pi$, where

$$M_3(u) = \begin{cases} (u+3\pi)^2/2 & -3\pi \leq u \leq -\pi \\ [(u+3\pi)^2-3(u+\pi)^2]/2 & -\pi \leq u \leq \pi \\ (3\pi-u)^2/2 & \pi \leq u \leq 3\pi. \end{cases}$$

In general, $Q_{r,s}(u)$ will be a combination of the derivatives of the r^{th} B-spline M_r.

References

1 Goldberg, R.R., Restrictions of Fourier transforms and extension of Fourier sequences, J. Approximation Theory, 3 (1970), 149-155.

2 Lee, S.L., Fourier transforms of B-splines and fundamental splines for cardinal Hermite interpolation, Proc. Amer. Math. Soc. (to appear).

3 Lipow, P.R. and I.J. Schoenberg, Cardinal interpolation and spline functions. III. Cardinal Hermite interpolation, J. Linear Algebra and Applic., 6 (1973), 273-304.

4 Marsden, M.J., F.B. Richards and S.D. Riemenschneider, Cardinal spline interpolation operators on ℓ^p data, Indiana Univ. Math. J., 24 (1975), 677-689.

5 Schoenberg, I.J., Notes on spline functions III: On the convergence of the interpolating cardinal splines as their degree tends to infinity, Israel J. Math., 16 (1973), 87-93.

6 Riemenschneider, S.D., Convergence of interpolating cardinal splines: Power growth, (to appear).

M.J. Marsden
Department of Mathematics
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

S.D. Riemenschneider*
Department of Mathematics
University of Alberta
Edmonton, Canada T6G 2G1

*Supported in part by Canadian National Research Council Grant A-7687.

n-WIDTH UNDER RESTRICTED APPROXIMATIONS

Avraham A. Melkman

The purpose is to find optimal approximating subspaces to a set of functions, when restrictions are placed on the approximation. Two instances are outlined. The first requires the interpolation of the function at a fixed set of points and may be connected to problems of optimal recovery. The second concerns monotone approximations, the main result involving the case when polynomials are optimal.

In the usual n-width problem (in the sense of Kolmogorov) one is given a subset A of Banach space X and the problem is to identify an optimal n-dimensional subspace X_n^* of X from which to approximate A, cf Lorentz [2]. Thus X_n^* achieves

$$d_n(A) = \inf_{X_n} \sup_{x\in A} \inf_{y\in X_n} \|x-y\|.$$

Since it is often desirable to have the approximation $y \in X_n$ satisfy some side conditions, it is natural to consider corresponding n-widths.

1. Interpolatory conditions.

This section is based on joint work with C. A. Micchelli. Take $X = L_2[0,1]$, $A = \bar{W}_2^r = \{f \mid f \in W_2^r[0,1],\ \|f^{(r)}\|_2 \le 1\}$ and let T_N be a fixed set of points $0 \le t_1 <\ldots< t_N \le 1$. Given $f \in \bar{W}_2^r$ and X_n we now consider only those approximations $y \in X_n$ which interpolate f at T_N, i.e. $y(t_i)=f(t_i), i=1,\ldots,N$,

and denote the corresponding, interpolatory, n-width by $d_n(\bar{W}_2^r;T_N)$.

THEOREM 1.1. *Let* $\mu_i(T_N)$ *and* $v_i(x)$ *be the eigenvalues and eigenvectors of*

$$v^{(2r)}(x) = (-1)^r \mu\, v(x) + \sum_{i=1}^{N} a_i\, \delta(x-t_i)$$

$$v^{(j)}(0) = v^{(j)}(1) = 0,\ j = r,\dots,2r-1;\ v(t_i)=0,\ i=1,\dots,N$$

a_i *free. Then* $d_n(\bar{W}_2^r;T_N) = \mu_n(T_N)^{-\frac{1}{2}}$.
Furthermore, let $\xi_1,\dots,\xi_n$ *and* $\eta_1,\dots,\eta_{n-r}$ *be the (simple) zeros of* $v_n(x)$ *and* $v_n^{(r)}(x)$ *respectively,* $n \geq N$. *Then optimal* n-*dimensional subspaces are*

i. *natural splines of degree* $2r-1$ *with knots at* $\xi_1,\dots,\xi_n$

ii. *ordinary splines of degree* $r-1$ *with knots at* $\eta_1,\dots,\eta_{n-r}$

In both cases $d_n(\bar{W}_2^r;T_N)$ *is achieved by interpolation at* $\xi_1,\dots,\xi_n$.

Conversely these results yield as a byproduct tighter L_2-bounds on the error in natural spline interpolation, by explicit estimation of $\mu_N(T_N)$. For example in cubic natural spline interpolation one obtains a bound of $(\Delta/\pi)^2\|f^{(2)}\|_2$, halving Schultz's estimate [5].

It is natural to ask which T_N^* minimizes $d_N(\bar{W}_2^r;T_N)$. Clearly a lower bound for the latter is $d_N(\bar{W}_2^r)$.

THEOREM 1.2. *Let* $\lambda_N, u_N(x)$ *be the* N-th *eigenvalue and eigenfunction of* $u^{(2r)}(x) = (-1)^r u(x)$, $u^{(j)}(0) = u^{(j)}(1) = 0$, $j=r,\dots,2r-1$, *characterizing the unrestricted* N-*width. Then* T_N^*

<u>is the set of zeros of</u> $u_N(x)$ <u>and</u> $\mu_N(T_N^*) = \lambda_N$.

COROLLARY 1.1. <u>The spline spaces of theorem 1.1 with</u> T_N^* <u>are optimal subspaces for the unrestricted</u> N-<u>width problem, in addition to the classical eigenfunction subspace. For example</u>: cubic natural spline interpolation at equally spaced points is an optimal approximation procedure for $\bar{W}_2^3$. This gives a concrete example of nonuniqueness, an issue discussed by Karlovitz [1].

These results may also be applied to the problem of L_2-optimal recovery of a smooth function from a sampling of its values at T_N, along the lines of Micchelli, Rivlin and Winograd [4]. Briefly, the interpolation procedures of theorem 1.1 provide optimal recovery schemes; the optimal choice for an additional sampling point is at the additional zero of $v_{N+1}(x)$; and, the most information is gained by putting the sampling points at the zeros of the function which is hardest to recover.

2. <u>Monotone approximation of monotone functions</u>.

Take $X = L_\infty[0,1]$, $A = M^r \equiv \{f \mid f \in \bar{W}_\infty^r,\ f' \geq 0\}$ and restrict the approximations to be monotone, denoting the corresponding n-width by $d_n^{\,+}(M^r)$.

THEOREM 2.1. $d_n^{\,+}(M^1) = \frac{1}{2} d_n(\bar{W}_\infty^1), d_n^{\,+}(M^2) = d_n(\bar{W}_\infty^2)$

$d_n(\bar{W}_\infty^r) < d_n^{\,+}(M^r) < c_r\, d_n(\bar{W}_\infty^r)$ <u>for some</u> c_r, $r > 2$.

The proof uses the observation that the unrestricted n-widths of M^r and $\bar{W}_\infty^r$ are the same. Further, for $r = 1,2$

the optimal approximation procedure for $\bar{W}^r_\infty$ is one of piece-wise linear interpolation at equally spaced points which for M^r yields a monotone interpolant. For $r > 2$ the upper bound is a restatement of DeVore's result (these proceedings) that the polynomial Jackson's theorems remain valid even with the mono-tinicity constraint. However, this yields no knowledge of the "worst approximable functions" or exact values of d^*_N. The int-erpolatory method, so successful for $\bar{W}^r_\infty$, cannot work here for it cannot ensure the monotinicity of the interpolant. It seems best, rather than good, approximations have to be used, a form-idable task. We therefore limit ourselves here to some results concerning $d^+_r(M^r)$ when the optimal subspace has to consist of polynomials of degree $r-1$

$$d^+_r(M^r) = \sup_{f \in M^r} \inf_{P'_{r-1} \geq 0} \|f - p_{r-1}\|_\infty .$$

Lorent and Zeller [3] showed that p^*_{r-1} is a best monotone approximation to f if and only if there exist a_i, $x_i, i=1,\dots,m$ and b_i, y_i, $i=1,\dots,\ell$ such that $|f(x_i) - p^*_{r-1}(x_i)| = \|f - p^*_{r-1}\|$, $p^{*\prime}_{r-1}(y_i) = 0$, $b_i > 0$, $m + \ell \leq r + 1$ and

(1) $$\sum_{i=1}^{m} a_i\, q(x_i) + \sum_{i=1}^{\ell} b_i\, q'(y_i) = 0$$

for any q a polynomial of degree $r-1$. This condition together with a consideration of duality can be used to prove a result stating roughly that only functions with many flat spots are candidates for the worst function.

LEMMA 2.1. $f \in M^r$ _can be a worst approximable function only if_

$Q = f - p^*_{r-1}$ is a perfect equioscillating spline of degree r with $m + 2\ell - r - 1$ knots, satisfying $Q'(y_i) = Q''(y_i) = 0$, $i = 1,\dots,\ell$, $Q(x_i) = (-1)^i \|Q\|$, $i = 1,\dots,m$.

The value $\|Q\|$ is a local maximum in the sense that any function close to f can be approximated at least as well. There are therefore many local maxima, corresponding to the possible values of m and ℓ, cf. Lorentz and Zeller [3], and the problem is to identify the global one. Intuitively one expects the worst function to be such that its best approximation is hampered most by the monotinicity constraint. Indeed

LEMMA 2.2. *If f is such that its best monotone approximation has $2\ell < r - 2$ then there exists a function which is worse approximable.*

These results lead to the following characterization.

THEOREM 2.2. *For r even, $d^+_r(M^r) = \|S_r\|_\infty$ where S_r is a polynomial of degree r equioscillating 3 times at $0, x_1, 1$, with $S_r'(y_i) = S_r''(y_i) = 0$, $i=1,\dots,(r-2)/2$, $S_r^{(r)} = 1$, and such that $0, x_1, 1$ and $y_i, i=1,\dots,(r-2)/2$ support a formula (1).*

The existence of such an S_r is ensured by theorem 16 of Lorentz and Zeller [3], and it is unique up to a reflection. For r odd a similar result holds with an additional condition $S_r'(0) = 0$.

References

1 Karlovitz L.A., Remarks on variational characterizations of eigenvalues and n-width problems, manuscript 1975, 18 pp.

2 Lorentz G.G., Approximation of functions, Holt, New York, 1966.

3 Lorentz G.G. and K.L. Zeller, Monotone approximation by algebraic polynomials, Trans.Amer.Math.Soc. 149(1970),1-18.

4 Micchelli C.A., T. J. Rivlin, and S. Winograd, The optimal recovery of smooth functions, to appear in Numer.Math.

5 Schultz, M.H., Error bounds for polynomial spline interpolation, Math. Comp. 24(1970), 507-515.

Avraham A. Melkman
Institute of Mathematics
The Hebrew University
Jerusalem, Israel.

THE EXACT ASYMPTOTIC VALUE FOR THE N-WIDTH OF SMOOTH FUNCTIONS IN L^∞

Charles A. Micchelli and Allan Pinkus

In this paper we answer a question raised by Chui and Smith and obtain the exact asymptotic value for the N-width of the set $D_r = \{f: ||Lf||_\infty \le 1,\ f \in W_\infty^r\,[0,1]\}$ where L is an r-th order totally disconjugate differential operator and $||\cdot||_\infty$ = sup norm on $[0,1]$.

1. Introduction

Let $W_\infty^r[0,1] = \{f: f^{(r-1)}$ absolutely continuous on $[0,1]$, $f^{(r)} \in L^\infty[0,1]\}$, $\lambda_j \in C^{r-j}[0,1]$, $j=1,\ldots,r$, and $(Lf)(x) = \prod_{j=1}^{r} (\frac{d}{dx} + \lambda_j(x))f(x)$. The N-width of the set

(1) $D_r = \{f: f \in W_\infty^r[0,1],\ ||Lf||_\infty \le 1\}$

(relative to $C[0,1]$) is defined by

(2) $d_N(D_r) = \inf_{X_N} \sup_{f \in D_r} \inf_{g \in X_N} ||f-g||_\infty$,

where the infimum is taken over all N-dimensional linear subspaces X_N of $C[0,1]$.

The purpose of this paper is to prove the following theorem which answers a question raised in Chui and Smith [1].

Let

$$e_r = \frac{2}{\pi^{r+1}} \sum_{-\infty}^{+\infty} \frac{(-1)^{k(r+1)}}{(2k+1)^{r+1}} .$$

THEOREM. For any $r \geq 1$,

$$\lim_{N\to\infty} N^r d_N(D_r) = e_r$$

The proof of the above theorem relies on several results of [3] which we summarize below.

2. Proof of theorem

Let $K(x,y)$ be the Green's function for the initial value problem

$$(Lf)(x) = h(x)$$
$$f^{(i)}(0) = 0,\ i=0,1,\ldots,r-1.$$

Also, define $k_0(x),\ldots,k_{r-1}(x)$ as the unique set of functions in the null space of L satisfying the conditions

$$k_i^{(j)}(0) = \delta_{ij},\ i,j=0,1,\ldots,r-1.$$

The Green's function $K(x,y)$ has the property that

$$K(x,y) = \begin{cases} 0 & , x < y \\ H(x,y), & x > y \end{cases}$$

where for each fixed y, $H(x,y)$ is in the null space of L, as a function of x.

Thus D_r has the equivalent representation

$$(3)\quad D_r = \{\sum_{j=0}^{r-1} a_j k_j(x) + \int_0^1 K(x,y)h(y)dy\colon \|h\|_\infty \leq 1,$$
$$(a_0,\ldots,a_{r-1}) \in R^r\}.$$

S. Karlin proves in [2] that for every integer $s \geq 0$ there exists a function

(4) $$P_s(x) = \sum_{j=0}^{r-1} b_j k_j(x) + \sum_{j=0}^{s} (-1)^j \int_{\xi_j}^{\xi_{j+1}} K(x,y)dy$$

$$0 = \xi_0 < \xi_1 < \ldots < \xi_s < \xi_{s+1} = 1$$

which equioscillates $r+s+1$ times, that is,

$$P_s(\tau_i) = (-1)^{i+1}||P_s||_\infty, \quad i = 1,\ldots,r+s+1,$$

for some points $0 \le \tau_1 < \tau_2 < \ldots < \tau_{r+s+1} \le 1$ (see also [3]). We will denote by Q_s the class of all functions which may be expressed as

$$P(x) = \sum_{j=0}^{r-1} a_j k_j(x) + \sum_{j=0}^{\ell} (-1)^j \int_{\eta_j}^{\eta_{j+1}} K(x,y)dy$$

for some constants $(a_0,\ldots,a_{r-1}) \in R^r$ and points $0 = \eta_0 < \eta_1 < \ldots < \eta_\ell < \eta_{\ell+1} = 1$, $\ell \le s$. Then P_s has the following properties:

(5) $||P_s||_\infty \le ||P||_\infty$, $P \in Q_s$,

and

(6) $\min\limits_{1\le j\le r+s+1} |f(x_j)| \le ||P_s||_\infty$,

where f is any function in D_r such that for some points $0 \le x_1 < \ldots < x_{r+s+1} \le 1$, $f(x_i)f(x_{i+1}) \le 0$, $i=1,\ldots,r+s$.

The importance of the function P_s rests on the equation

(7) $d_N(D_r) = ||P_{N-r}||_\infty$, $N \ge r$,

which, along with (5) and (6), was proven in [3].

We are now prepared to prove the theorem.

PROOF. For every integer N, let

$$G_N(x) = \sum_{j=0}^{N-1} (-1)^j \int_{\frac{j}{N}}^{\frac{j+1}{N}} K(x,y)dy.$$

We claim that there exists a v_N in the null space of L such that for the function $H_N = G_N + v_N$

$$(8) \quad \lim_{N\to\infty} N^r \|H_N\|_\infty = e_r$$

and there exist N point $0 \le x_1^N < \ldots < x_N^N \le 1$ such that

$$(9) \quad \lim_{N\to\infty} \max_{1\le i\le N} \left| N^r H_N(x_i^N) - (-1)^{i+1} e_r \right| = 0.$$

These facts, together with (5), (6), and (7) imply

$$\min_{1\le i\le N+1} |H_{N+1}(x_i^{N+1})| \le d_N(D_r) \le \|H_{N-r+1}\|_\infty .$$

Thus we conclude the validity of the theorem.

Let us then prove (8) and (9). Recall that the r-th Euler polynomial is defined by the relation

$$E_r(x+1) + E_r(x) = \frac{2x^r}{r!}, \quad x \in R.$$

Here we have normalized E_r so that $E_r^{(r)}(x) = 1$. We perform the usual surgery on E_r and define $\overline{E}_r$ by

$$(10) \quad \begin{aligned} \overline{E}_r(x) &= E_r(x), \quad x \in [0,1] \\ \overline{E}_r(x+1) &= -\overline{E}_r(x), \quad x \in R. \end{aligned}$$

It is evident that we may express G_N as

$$G_N(x) = \frac{1}{N^r} \int_0^1 \left(\frac{d^r}{dy^r} \overline{E}_r(Ny)\right) K(x,y)dy.$$

Our next step is to integrate the above expression by parts. For any $f \in W_\infty^r[0,1]$, there exists constants $c_0, c_1, \ldots, c_{r-1}$ such that

$$\int_0^1 f^{(r)}(y)K(x,y)dy = f(x) - \sum_{j=0}^{r-1} c_j k_j(x) + \int_0^x J(x,y)f(y)dy$$

where

$$J(x,y) = (-1)^r \frac{\partial^r}{\partial y^r} K(x,y), \quad y < x.$$

Applying this identity to $G_N(x)$ it follows that there exists v_N in the null space of L such that

$$G_N(x) = \frac{1}{N^r} \overline{E}_r(Nx) - v_N(x) + \frac{1}{N^r} \int_0^x J(x,y)\overline{E}_r(Ny)dy.$$

We define $H_N = G_N + v_N$. From (10) we note that (8) and (9) will follow provided that

(11) $||E_r||_\infty = e_r$

and

(12) $\lim_{N\to\infty} \max_{0\le x\le 1} \left|\int_0^x J(x,y)\overline{E}_r(Ny)dy\right| = 0.$

The expression (11) for the L^∞-norm of E_r is well-known. It is easily deduced from the Fourier series expansion of $e^{i\pi x}E_r(x)$ and the fact that

$$||E_r||_\infty = \begin{cases} |E_r(\frac{1}{2})| \ , & r \text{ even} \\ |E_r(0)| \ , & r \text{ odd} \end{cases}$$

Thus it remains to verify (12).

Let M be an integer. Divide $[0,1]$ into M equal pieces, $I_i = [\frac{i}{M}, \frac{i+1}{M}]$, $i = 0,1,\ldots,M-1$. Let $g_i(y)$ be the characteristic function of the interval I_i and define $S_M(x,y) = \sum_{i=0}^{M-1} J(x,\frac{i}{M}) g_i(y)$. Then $\lim_{M\to\infty} ||J-S_M||_\infty = 0$, where $||\cdot||_\infty$ denotes the L^∞-norm on $[0,1] \times [0,1]$. Now, for any

$x \in [0,1]$

$$\left|\int_0^x J(x,y)\overline{E}_r\,(Ny)\,dy\right|$$

$$\leq \|J-S_M\|_\infty \int_0^1 |\overline{E}_r(Ny)|\,dy + M\|J\|_\infty \max_{0\leq i\leq M+1} \left|\int_{I_i} \overline{E}_r(Ny)\,dy\right|.$$

However, $\left|\int_{I_i} \overline{E}_r(Ny)\,dy\right| \leq \frac{1}{N}\int_0^1 |E_r(y)|\,dy$ for all $i=0,1,\ldots,M-1$.

Thus

$$\lim_{N\to\infty} \max_{0\leq x\leq 1} \left|\int_0^x J(x,v)\overline{E}_r(Ny)\,dy\right| \leq \|J-S_M\|_\infty \int_0^1 |E_r(y)|\,dy.$$

Letting $M\to\infty$ we obtain (12) and thus the proof is complete.

REMARK. We conjecture that the above theorem remains valid for any r-th order differential operator

$$L = D^r + \sum_{j=0}^{r-1} a_j D^j.$$

Our proof, however, requires the upper and lower bounds given by (5) and (6) which were proven in [3] only for differential operators L allowing a global factorization into real linear factors.

References

1 Chui, C.K. and P.W. Smith, Some nonlinear apline approximation problems related to N-widths, J. Approximation Theory, 13 (1975), 421-430.

2 Karlin, S., Oscillatory perfect splines and related extremal problems, in Studies in Spline Functions and Approximation Theory, ed. by S. Karlin, C. A. Micchelli, A. Pinkus and I. J. Schoenberg, Academic Press, New York, 1976.

3 Micchelli, C.A. and A. Pinkus, On N-widths in L^∞, IBM Research Report RC 5478 (1975).

Charles A. Micchelli
Mathematical Sciences Department
IBM T. J. Watson Research Center
Yorktown Heights, N.Y. 10598

Allan Pinkus
Mathematics Research Center
University of Wisconsin
Madison, Wisconsin 53706

ON n-WIDTHS AND OPTIMAL RECOVERY IN $\mathcal{M}^r$

Charles A. Micchelli and Allan Pinkus

Let BV[0,1] denote the set of all functions of bounded variation on [0,1]. For any $\lambda \in$ BV[0,1], we set $\|\lambda\|$ = total variation of λ, and define

$$\mathcal{M}^r = \mathcal{M}^r[0,1] = \{f : f(x) = \sum_{i=0}^{r-1} a_i x^i + \frac{1}{(r-1)!}\int_0^1 (x-t)_+^{r-1} d\lambda(t),$$
$$\lambda \in BV[0,1]\} .$$

In this paper, the n-width both in the sense of Kolmogorov and Gel'fand, for

$$\mathcal{B}_r = \{f : f \in \mathcal{M}^r, \|\lambda\| \le 1\}$$

is found. In addition, we solve the related problem of optimal recovery of a function $f \in \mathcal{B}_r$.

1 n-Widths

Let $d_n(\mathcal{B}_r; L^1[0,1]) = d_n(\mathcal{B}_r)$ denote the n-width in the sense of Kolmogorov of $\mathcal{B}_r$ in $L^1[0,1]$, given explicitly by

$$d_n(\mathcal{B}_r) = \inf_{X_n} \sup_{f \in \mathcal{B}_r} \inf_{g \in X_n} \|f - g\|_1,$$

where the infimum is taken over all n-dimensional linear subspaces X_n of $L^1[0,1]$ and $\|f\|_1 = \int_0^1 |f(t)|\, dt$. The n-width in the sense of Gel'fand for $\mathcal{B}_r$ in C[0,1] is given as

$$d^n(\mathcal{B}_r) = d^n(\mathcal{B}_r; C[0,1]) = \inf_{L_n} \sup_{f \in \mathcal{B}_r \cap L_n} \|f\|_1,$$

where the infimum is taken over all subspaces L_n of C[0,1]

of codimension n. An extremal subspace for either $d_n(\mathcal{B}_r)$ or $d^n(\mathcal{B}_r)$ is any subspace for which the respective infimum is attained.

A perfect spline P of degree r with s knots at $x_1, \ldots, x_s$, $0 = x_0 < x_1 < \ldots < x_s < x_{s+1} = 1$ is any function expressible as

$$P(x) = \sum_{i=0}^{r-1} a_i x^i + c \sum_{j=0}^{s} (-1)^j \int_{x_j}^{x_{j+1}} (x-t)_+^{r-1} dt.$$

The following result is essentially contained in Karlin [1].

THEOREM 1. <u>For</u> $n \geq r$, <u>there exists a perfect spline</u> $Q_{n,r} = Q$, <u>unique up to multiplication by</u> -1, <u>of degree</u> r <u>with</u> n <u>knots such that</u> $\|Q^{(r)}\|_\infty = 1$ <u>which satisfies the boundary conditions</u> $Q^{(i)}(0) = Q^{(i)}(1) = 0$, $i = 0, 1, \ldots, r-1$, <u>and equioscillates at</u> $n - r + 1$ <u>points in</u> $(0,1)$, i.e., $Q(\eta_i) = (-1)^i \sigma \|Q\|_\infty$, $i = 1, 2, \ldots, n - r + 1$, $0 < \eta_1 < \ldots < \eta_{n-r+1} < 1$, $\sigma = +1$ <u>or</u> -1, <u>fixed</u>.

Utilizing Theorem 1, we obtain the following result

THEOREM 2. <u>For</u> $r \geq 2$,

(1) $$d_n(\mathcal{B}_r) = d^n(\mathcal{B}_r) = \begin{cases} \infty, & n < r \\ \|Q_{n,r}\|_\infty, & n \geq r \end{cases}$$

(2) $Q_{n,r}$ <u>has exactly</u> $n - r$ <u>simple zeros</u> $\{\zeta_i\}_{i=1}^{n-r}$ <u>in</u> $(0,1)$, <u>and the linear space spanned by the functions</u> $\{1, x, \ldots, x^{r-1}, (x-\zeta_1)_+^{r-1}, \ldots, (x-\zeta_{n-r})_+^{r-1}\}$ <u>is an extremal subspace for</u> $d_n(\mathcal{B}_r)$.

(3) $L_n^* = \{g : g \in C[0,1],\ g(\xi_i) = 0,\ i = 1, 2, \ldots, n\}$, <u>where</u> $\{\xi_i\}_{i=1}^n$ <u>are the</u> n <u>knots of</u> $Q_{n,r}$ <u>is an extremal subspace for</u> $d^n(\mathcal{B}_r)$.

The proof of Theorem 2 is based upon duality and analysis similar to that found in [2].

2 Optimal Recovery

Let $\underline{x} = (x_1, \ldots, x_n)$, $x_0 = 0 < x_1 < \ldots < x_n < 1 = x_{n+1}$, be given and denote the vector $(f(x_1), \ldots, f(x_n))$ by $\underline{f}$. The problem of optimal recovery is one of determining a rule for best recovering $f \in \mathcal{B}_r$, $n \geq r \geq 2$, based on the information $\underline{f}$ (for the L^∞ analogue of this problem see [3]).

Any transformation R from $\{\underline{f}: f \in \mathcal{B}_r\}$ into $L^1[0,1]$ determines a recovery scheme $Sf = R\underline{f}$ for $f \in \mathcal{B}_r$. The error for recovery based on S is defined as

$$\|I - S\|_1 = \sup_{f\in\mathcal{B}_r} \|f - Sf\|_1,$$

and

$$E(\underline{x}) = \inf\{\|I - S\|_1 : S \text{ a recovery scheme}\}$$

is the minimum error in recovering $f \in \mathcal{B}_r$ from the information $\underline{f}$. S^* is called an optimal recovery scheme provided that $\|I - S^*\|_1 = E(\underline{x})$.

THEOREM 3. <u>For every</u> $n \geq r \geq 2$, <u>and each</u> $\underline{x}$ <u>as above, there exists a function</u>

$$P_{\underline{x}}(x) = \sum_{j=0}^{n} \frac{(-1)^j}{(r-1)!} \int_{x_j}^{x_{j+1}} (x - t)_+^{r-1} dt + \sum_{j=1}^{n} c_j (x - x_j)_+^{r-1}$$

<u>which satisfies the boundary conditions</u> $P_{\underline{x}}^{(i)}(1) = 0$, $i = 0, 1, \ldots, r-1$, <u>and equioscillates</u> $n - r + 1$ <u>times on</u> $(0,1)$. $P_{\underline{x}}$ <u>has</u> $n - r$ <u>simple zeros</u> $\{\zeta_i(\underline{x})\}_{i=1}^{n-r}$, $0 < \zeta_1(\underline{x}) < \ldots < \zeta_{n-r}(\underline{x}) < 1$, <u>and</u> $x_i < \zeta_i(\underline{x}) < x_{i+r}$, $i = 1, \ldots, n - r$.

Based on Theorem 3, we define a recovery scheme S^* by interpolating the function f at the values $\{x_i\}_{i=1}^n$ by linear combinations of $1, x, \ldots, x^{r-1}, (x - \zeta_1(\underline{x}))_+^{r-1}, \ldots,$

$(x - \zeta_{n-r}(\underline{x}))_+^{r-1}$, i.e., $(S^*f)(x_i) = f(x_i)$, $i = 1, \ldots, n$. Since $x_i < \zeta_i(\underline{x}) < x_{i+r}$, $i = 1, \ldots, n - r$, such a recovery scheme is well-defined.

We prove the following theorem by using Theorems 1-3, and analysis paralleling that used in the proof of Theorem 2.

THEOREM 4. S^* is an optimal recovery scheme for $\mathcal{B}_r$, and $E(\underline{x}) = \|P_{\underline{x}}\|_\infty$. Furthermore, if $\underline{\xi} = (\xi_1, \ldots, \xi_n)$ are the knots of $Q_{n,r}$, then $\min_{\underline{x}} E(\underline{x}) = E(\underline{\xi}) = \|Q_{n,r}\|_\infty = \min_{\underline{x}}\|P_{\underline{x}}\|_\infty$.

Full details and extensions of all the above results will appear elsewhere.

References

1 Karlin, S., Oscillatory Perfect Splines and Related Extremal Problems, to appear in Studies in Spline Functions and Approximation Theory, ed. S. Karlin, C. A. Micchelli, A. Pinkus, I. J. Schoenberg, Academic Press.

2 Micchelli, C. A., and A. Pinkus, On n-Widths in L^∞, IBM Research Report 5478 (1975).

3 Micchelli, C. A., T. J. Rivlin, and S. Winograd, The Optimal Recovery of Smooth Functions, to appear in Numer. Math.

Charles A. Micchelli
IBM T. J. Watson Res. Ctr.
Math. Sc. Dept.
Yorktown Hts., N.Y. 10598

Allan Pinkus*
Mathematics Research Center
University of Wisconsin
Madison, Wis. 53705

*Sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and the National Science Foundation under Grant No. MPS72-00381A01.

ON NIKOLSKII-TYPE INEQUALITIES FOR

ORTHOGONAL EXPANSIONS

R. J. Nessel and G. Wilmes

Nikolskii-type inequalities, i.e. inequalities between different norms of a function, are established for polynomials $\Sigma_{k=0}^{n} c_k \Phi_k$ where $\{\Phi_k\}$ is a regular orthonormal system in weighted Lebesgue spaces L_w^p, $1 \leq p \leq \infty$. As applications, expansions into Legendre polynomials and Hermite functions are considered.

Nikolskii-type inequalities, following the usual nomenclature, are inequalities between different norms of the same function. Thus let us consider trigonometric polynomials $t_n(x): = \Sigma_{k=-n}^{n} c_k e^{ikx}$ in $L_{2\pi}^p$ with

$$\|f\|_{p,2\pi}: = \begin{cases} [(1/2\pi)\int_{-\pi}^{\pi} |f(u)|^p du]^{1/p}, & 1 \leq p < \infty \\ \text{ess. sup } |f(u)| & , p = \infty, \end{cases}$$

respectively (we only consider these values of p though there are of course counterparts for any $p > 0$). Then S. M. Nikolskii (1951) proved that

$$(1) \quad \|t_n\|_{q,2\pi} \leq C_{p,q}\, n^{1/p - 1/q} \|t_n\|_{p,2\pi} \quad (1 \leq p \leq q \leq \infty)$$

with $C_{p,q}$ independent of n. Let us mention that inequalities of this type were already considered by D. Jackson (1933) in the particular case $q = \infty$; the result (1) itself was obtained independently also by G. Szegö-A. Zygmund (1953).

The example of the Fejér kernel shows (cf. [7, p.230]) that the order in (1) is the correct one. The best result

known so far with respect to the constant reads

$$\|t_n\|_{q,2\pi} \leq (2p_o n+1)^{1/p - 1/q}\|t_n\|_{p,2\pi} \quad (1 \leq p \leq q \leq \infty),$$

p_o denoting the smallest integer not less than $p/2$. For this one may consult papers of I. I. Ibragimov (1959) as well as A. F. Timan [7, p. 227ff]. The exact value of the constant in (1), however, seems to be an open problem. Corresponding results are available for entire functions of exponential type as well as for functions of several variables. But for detailed bibliographical comments as well as for certain extensions we refer to [2;4].

Nikolskii inequalities have received considerable attention because of their wide applicability in various problems of analysis. Thus, for example, Jackson was interested in deriving L^∞-estimates from known L^2-results in connection with certain orthogonal systems; Nikolskii used (1) to obtain embedding theorems for his spaces. So it seems to be of some interest to treat the matter in the frame of arbitrary orthogonal expansions. To this end, let us consider the following concrete situation.

Let $L^p_w(a,b)$, $1 \leq p \leq \infty$, $-\infty \leq a < b \leq \infty$, be the space of functions, pth power integrable with respect to the weight $w(x) \geq 0$:

$$\|f\|_{p,w} := \begin{cases} (\int_a^b |f(u)|^p w(u)du)^{1/p}, & 1 \leq p < \infty \\ \text{ess. sup } |f(u)| & , p = \infty . \end{cases}$$

Let $\{\Phi_k\}_{k=0}^\infty$ be an orthonormal system such that $\{\Phi_k\} \subset L^p_w$ for all $1 \leq p \leq \infty$, and suppose that there exist a constant $M > 0$ and some real $\delta \in \mathbb{R}$ such that

$$(2) \quad \|\Phi_o\|_{\infty,w} \leq M, \quad \|\Phi_k\|_{\infty,w} \leq Mk^\delta, \quad k > 0.$$

Then to $f \in L^p_w(a,b)$ one may associate its Fourier series

$$f \sim \Sigma_{k=0}^{\infty} f^{\wedge}(k)\Phi_k, \quad f^{\wedge}(k) := \int_a^b f(u)\overline{\Phi_k(u)}w(u)du.$$

Let us consider the Cesàro means of order $j \in \mathbb{P}$ (the set of nonnegative integers)

$$(C,j)_n f := \Sigma_{k=0}^{n}(A^j_{n-k}/A^j_n)f^{\wedge}(k)\Phi_k, \quad A^j_n := \binom{n+j}{n}.$$

Then, following Stein [6], the system $\{\Phi_k\}$ is said to be regular if for some $j \in \mathbb{P}$ there exists a constant $A > 0$ such that for all $1 \leq p \leq \infty$, $f \in L^p_w(a,b)$, $n \in \mathbb{P}$

$$(3) \quad \|(C,j)_n f\|_{p,w} \leq A\|f\|_{p,w}.$$

Denoting by Π_n the set of all polynomials $r_n := \Sigma_{k=0}^{n} c_k\Phi_k$ of degree n, we have

THEOREM. Let the system $\{\Phi_k\}$ be regular (for some $j \in \mathbb{P}$), and suppose that (2) holds true for some $\delta \in \mathbb{R}$. Then

$$(4) \quad \|r_n\|_{q,w} \leq \left(C^p \begin{cases} n^{(2\delta+1)}, & \delta > -1/2 \\ \log n, & \delta = -1/2 \\ 1, & \delta < -1/2 \end{cases} \right)^{1/p-1/q} \|r_n\|_{p,w}$$

for any $1 \leq p \leq q \leq \infty$ with C independent of n, p, q.

Proof. Let us consider, for example, the case $\delta > -1/2$. We essentially follow the device of Stein [6] in how to use the Riesz-Thorin theorem for operators given on certain linear subspaces such as the identity operator on the set of polynomials. Indeed, it is an immediate consequence of (2) that

$$(5)\quad \|r_n\|_{\infty,w} \le \|r_n\|_{1,w} \Sigma_{k=0}^{n} \|\Phi_k\|_{\infty,w}^2 \le C' M^2 n^{2\delta+1} \|r_n\|_{1,w}$$

for some absolute constant C'. Let λ be a smooth function on $[0,\infty)$ such that $\lambda(t) = 1$ for $0 \le t \le 1$ and $=0$ for $t \ge 2$. Then the means

$$L_n f := \Sigma_{k=0}^{\infty} \lambda(k/n) f^\wedge(k) \Phi_k$$

possess the following properties:

(6) (i) $L_n f \in \Pi_{2n}$ for all $f \in L_w^p(a,b)$,

(ii) $L_n r_n = r_n$ for all $r_n \in \Pi_n$,

(iii) $\|L_n f\|_{p,w} \le (A \int_0^2 t^j |\lambda^{(j+1)}(t)| dt) \|f\|_{p,w}$ uniformly for all $1 \le p \le \infty$, $f \in L_w^p(a,b)$, and $n \in \mathbb{P}$,

(iv) $\|L_n f\|_{\infty,w} \le C'' M^2 n^{2\delta+1} \|f\|_{1,w}$ for all $f \in L_w^1(a,b)$.

Indeed, (i), (ii) are obvious; (iii) follows by a multiplier criterion as given in [1], so that (iv) is an immediate consequence of (i), (5), and (iii). Thus, the bounded linear operator L_n, well-defined on all $L_w^p(a,b)$-spaces, $1 \le p \le \infty$, is of type $(1,1)$, $(1,\infty)$, (∞,∞), so that the classical Riesz-Thorin convexity theorem implies that L_n is of type (p,q) for all $1 \le p \le q \le \infty$. In view of the projection property (6,ii) this completes the proof of (4).

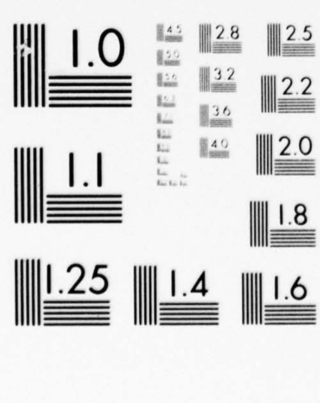
Obviously (4) reproduces the original result (1) of Nikolskii, apart from the constant. Concerning arbitrary orthonormal expansions, if (2) is satisfied for some $\delta > 0$, then the order in (4) is increased in comparison to the trigonometric case. But this cannot be removed in general. Indeed, let $a = -1$, $b = +1$, $w(x) \equiv 1$, and consider the (normalized) Legendre polynomials

AD-A033 032 TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS F/G 12/1
APPROXIMATION THEORY II
NOV 76 LORENTZ, G.; CHUI, C.; SCHUMAKER, L.
UNCLASSIFIED AFOSR TR-76-1242 N/L

6 OF 7
AD
A033032

OF 7

033032



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

$$\Phi_k^*(x) := (-1)^k[2^k k!]^{-1}\sqrt{k + 1/2}(d/dx)^k[(1 - x^2)^k].$$

Then $\|\Phi_k^*\|_\infty = \Phi_k^*(1) = \sqrt{k + 1/2}$ so that (2) holds true for $\delta = 1/2$. Since (3) is satisfied if e.g. $j = 1$, inequality (4) implies for any polynomial $r_n^* := \Sigma_{k=0}^n c_k \Phi_k^*$ that

$$\|r_n^*\|_q \leq C^{(1-p/q)} n^{2(1/p-1/q)} \|r_n^*\|_p \quad (1 \leq p \leq q \leq \infty).$$

Here the order cannot be improved. Indeed, if $P_n(x)$ is an arbitrary algebraic polynomial of degree n on the interval $[-1,1]$, then it is a well-known consequence of Markoff's inequality that

$$\|P_n\|_q \leq ((p + 1)n^2)^{1/p-1/q} \|P_n\|_p \quad (1 \leq p \leq q \leq \infty),$$

and this is best possible with respect to the order (cf. [7, p.236]).

On the other hand, M. F. Timan [8] announced examples of unbounded expansions (e.g. the Haar system) which nevertheless satisfy (4) with order $n^{1/p-1/q}$.

But there are also examples of expansions satisfying (2) for some negative δ. In fact, let $a = -\infty$, $b = +\infty$, $w(x) \equiv 1$, and consider the Hermite functions

$$\Phi_k^{**}(x) := (-1)^k(2^k k! \sqrt{\pi})^{-1/2} e^{x^2/2} (d/dx)^k e^{-x^2}.$$

Then (2) is satisfied for $\delta = -1/12$ (cf. [3, p. 571]). Since (3) holds true if e.g. $j = 1$, the Theorem yields for any polynomial r_n^{**}

$$\|r_n^{**}\|_q \leq C^{(1-p/q)} n^{(5/6)(1/p-1/q)} \|r_n^{**}\|_p \quad (1 \leq p \leq q \leq \infty).$$

Let us finally note that one may formulate analogs of the Theorem in the setting of abstract Banach spaces. In fact, the multiplier criterion used in the course of the proof works in connection with arbitrary spectral measures in Banach spaces satisfying suitable summability hypotheses. For this as well as for further extensions and details, however, see [5].

References

1 Butzer, P. L., R. J. Nessel, and W. Trebels, On summation processes of Fourier expansions in Banach spaces. II. Saturation theorems, Tôhoku Math. J., 24 (1972), 551-569.

2 Görlich, E. and G. Paulus, A bibliography on Bernstein-type inequalities (to appear).

3 Hille, E. and R. S. Phillips, Functional Analysis and Semi-Groups, Providence, R. I. 1957.

4 Nessel, R. J. and G. Wilmes, Nikolskii-type inequalities for trigonometric polynomials and entire functions of exponential type (to appear).

5 Nessel, R. J. and G. Wilmes, Inequalities of Bernstein-Nikolskii-type for regular spectral measures (to appear).

6 Stein, E. M., Interpolation in polynomial classes and Markoff's inequality, Duke Math. J. 24 (1957), 467-476.

7 Timan, A. F., Theory of Approximation of Functions of a Real Variable, Mac Millan, New York 1963.

8 Timan, M. F., On a property of orthonormed systems satisfying an inequality of S. M. Nikolskii (Russian), Abstracts, Kalouga 1975, 101-102.

R. J. Nessel, G. Wilmes*
Lehrstuhl A für Mathematik
RWTH Aachen
Templergraben 55
51 Aachen, BRD

*Supported by Grant No. II B7-FA 5844 awarded by the Minister für Wissenschaft und Forschung des Landes Nordrhein-Westfalen.

HYBRID FINITE ELEMENT METHODS

J. T. Oden

Some new techniques for determining error estimates for so-called hybrid finite-element methods for the approximation of solutions of linear elliptic boundary value problems of second-order are described.

1. Introduction

The hybrid finite element methods, developed largely by Pian and Tong and their associates (e.g. [1]) have been used successfully in the numerical solution of a variety of practical problems, and there appear to be certain classes of problems in which these methods have advantages over conventional finite element techniques (e.g. shell problems, elasto-plasticity problems, problems of cracks and stress singularities). However, the intrinsic mathematical properties of these methods have only recently begun to be investigated [2-4]. We describe here a fairly general theory of mixed-hybrid methods developed in [4].

2. A Mixed-Hybrid Variational Principle

Consider as a model problem,

$$(2.1) \qquad \begin{aligned} -\Delta u + u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

where Δ is the Laplacian operator in $\mathbb{R}^2$, $f \in L_2(\Omega)$, and Ω is a convex polygonal domain in $\mathbb{R}^2$.

Let P denote a triangulation of $\overline{\Omega}$ into E triangles $\overline{\Omega}_e$, i.e.,

(2.2) $$\bar{\Omega} = \bigcup_{e=1}^{E} \bar{\Omega}_e \ , \quad \Omega_e \cap \Omega_f = \emptyset \quad \text{if} \quad e \neq f .$$

We also define a collection of boundary pieces of P

(2.3) $$\Gamma = \bigcup_{e=1}^{E} \Gamma_e \ , \quad \Gamma_e = \partial\Omega_e - S_e \ , \quad \Gamma_{ef} = \Gamma_e \cap \Gamma_f$$

where S_e is the set of vertices of $\bar{\Omega}_e$.

We now introduce the following definitions and conventions:

I. Spaces Defined on the Partition P

(2.4) $$H^m(P) = \{u: u_e \in H^m(\Omega_e) \ , \quad 1 \leq e \leq E\}$$

(2.5) $$||u||_{H^m(P)} = [\sum_{e=1}^{E} ||u_e||^2_{H^m(\Omega_e)}]^{\frac{1}{2}}$$

(2.6) $$\mathcal{W}(\Gamma) = \text{completion of } L_2(\Gamma) \text{ in } ||\cdot||_{\mathcal{W}(\Gamma)}$$

where

(2.7) $$||\psi||_{\mathcal{W}(\Gamma)} = [\sum_{e=1}^{E} ||\psi_e||^2_{H^{-\frac{1}{2}}(\partial\Omega_e)}]^{\frac{1}{2}}$$

in which $||\cdot||_{H^{-\frac{1}{2}}(\partial\Omega_e)}$ is a dual norm of $||\cdot||_{H^{\frac{1}{2}}(\partial\Omega_e)}$.

II. Product Space

(2.8) $$\mathcal{X} = H^1(P) \times \underset{\sim}{L}_2(P) \times \mathcal{W}(\Gamma)$$

(2.9) $$||\underset{\sim}{\lambda}||_{\mathcal{X}} = [||u||^2_{H^1(P)} + ||\underset{\sim}{\sigma}||^2_{\underset{\sim}{L}_2(P)} + ||\psi||_{\mathcal{W}(\Gamma)}]^{\frac{1}{2}}$$

where $\underset{\sim}{L}_2(P) = L_2(P) \times L_2(P)$ and $\underset{\sim}{\sigma} = (\sigma_1, \sigma_2)$.

III. Special Bilinear and Linear Forms

(2.10) $$B: X \times X \to \mathbb{R}\ , \quad B(\underset{\sim}{\lambda},\underset{\sim}{\overline{\lambda}}) \sum_{e=1}^{E} b_e(\underset{\sim}{\lambda}_e,\underset{\sim}{\overline{\lambda}}_e)$$

$\underset{\sim}{\lambda},\underset{\sim}{\overline{\lambda}}$ = arbitrary triples in X; $\underset{\sim}{\lambda} = (u,\underset{\sim}{\sigma},\psi)$

$$b_e(\underset{\sim}{\lambda}_e,\overline{\lambda}_e) = \int_{\Omega_e} [\underset{\sim}{\sigma}_e \cdot \nabla\overline{u}_e + u_e\overline{u}_e + (\nabla u_e - \underset{\sim}{\sigma}_e) \cdot \underset{\sim}{\overline{\sigma}}_e]dx_1dx_2$$

(2.11) $$+ \oint_{\partial\Omega_e} (\psi_e\overline{u}_e + \overline{\psi}_e u_e)ds$$

(2.12) $$F(\underset{\sim}{\overline{\lambda}}) = \sum_{e=1}^{E} f_e(\underset{\sim}{\overline{\lambda}}_e)\ ; \quad f_e(\underset{\sim}{\overline{\lambda}}_e) = \int_{\Omega_e} f_e\overline{u}_e\ dx_1dx_2$$

IV. Properties of the Forms $B(\cdot,\cdot)$ and $F(\cdot)$

In [7], the forms $B(\underset{\sim}{\lambda},\underset{\sim}{\overline{\lambda}})$ and $F(\underset{\sim}{\overline{\lambda}})$ are shown to have certain properties which guarantee that the variational problem of finding $\underset{\sim}{\lambda} \in X$ such that

(2.13) $$B(\underset{\sim}{\lambda},\underset{\sim}{\overline{\lambda}}) = F(\underset{\sim}{\overline{\lambda}}) \qquad \forall\ \underset{\sim}{\overline{\lambda}} \in X$$

has a unique solution in X. Moreover, it can be shown that the mixed-hybrid variational problem (2.15) is equivalent to the weak form of (2.1).

3. Finite Element Approximations

The variational principle described in the previous section has been designed so as to be naturally adaptable to mixed-hybrid finite-element approximations. As expected, the triangles $\overline{\Omega}_e$ are now viewed as finite elements.

We introduce the spaces of polynomials: $Q_k^1(P) = \{U \in H^1(P),\ U_e \in P_k(\Omega_e),\ 1 \leq e \leq E\}$, $\underset{\sim}{Q}_r^o(P) = \{\underset{\sim}{\Sigma} \in \underset{\sim}{L}_2(P),\ \underset{\sim}{\Sigma}_e \in \underset{\sim}{P}_r(\Omega_e),\ 1 \leq e \leq E\}$, and $Q_t^{-\frac{1}{2}}(\Gamma) = \{\Psi \in W(\Gamma),\ \psi_{ef} = \psi|_{\Gamma_{ef}} \in P_t(\Gamma_{ef}),$

$1 \leq e,\ f \leq E,\ e > f\}$ where $P_k(\Omega_e)$ is the space of polynomials of degree $\leq k$ on Ω_e, etc. These spaces have the following interpolation properties: given $u \in H^{\ell}(\Omega_e)$, $\underset{\sim}{\sigma} \in (H^q(\Omega_e))^2$, $\psi \in \hat{H}^{m-3/2}(\Omega_e, \partial\Omega_e)$ (= completion of the space of normal derivatives $\partial v/\partial n \in L_2(\partial\Omega_e)$ in the norm $||\psi||_{\hat{H}^{m-3/2}(\partial\Omega_e)} = \inf\{||v||_{H^m(\Omega_e)};\ m \geq 1,\ \left.\frac{\partial v}{\partial n}\right|_{\partial\Omega_e} = \psi\}$) there exists constants C_1, C_2, $C_3 > 0$, <u>independent of the partition</u> $\underline{P}$, and elements U, $\underset{\sim}{\Sigma}$, and ψ in the respective spaces in (3.1), such that

$$
(3.1)\qquad
\begin{aligned}
||u - U||_{H^s(\Omega_e)} &\leq C_1 h_e^{\mu} ||u||_{H^{\ell}(\Omega_e)} \\
||\underset{\sim}{\sigma} - \underset{\sim}{\Sigma}||_{\underset{\sim}{L}_2(\Omega_e)} &\leq C_2 h_e^{\nu} ||\underset{\sim}{\sigma}||_{\underset{\sim}{H}^q(\Omega_e)} \\
||\psi - \Psi||_{H^{-\frac{1}{2}}(\partial\Omega_e)} &\leq C_3 h_e^{\theta} ||\psi||_{\hat{H}^{m-3/2}(\partial\Omega_e)}
\end{aligned}
$$

where $h_e = \mathrm{dia}(\Omega_e)$ and (with $0 \leq s \leq 1$, $m \geq 1$),

$$
\mu = \min(k+1-s,\ \ell-s);\quad \nu = \min(r+1, q);\quad \theta = \min(t+\tfrac{3}{2},\ m-1).
$$
(3.2)

Let $\underset{\sim}{\Lambda} = (U, \underset{\sim}{\Sigma}, \Psi) \in Q_k^1(P) \times \underset{\sim}{Q}_r^o(P) \times Q_t^{-\frac{1}{2}}(\Gamma) \equiv X_h \subset X$.

Then the mixed-hybrid finite element method consists of seeking the $\underset{\sim}{\Lambda} \in X_h$ such that

$$
(3.3)\qquad B(\underset{\sim}{\Lambda}, \overline{\underset{\sim}{\Lambda}}) = F(\overline{\underset{\sim}{\Lambda}}) \qquad \forall\ \overline{\underset{\sim}{\Lambda}} \in X_h\ ,
$$

where $B(\cdot,\cdot)$ and $F(\cdot)$ are the forms defined previously.

Let Π_e^1 and Π_e^0 denote orthogonal projections of $H^1(\Omega_e)$ and $L_e(\Omega_e)$ onto $Q_k^1(\Omega_e)$ and $Q_r^o(\Omega_e)$, respectively. We introduce the special stability parameters

$$
\begin{aligned}
\mu_e &= \inf_{\Psi \in Q_t^{-\frac{1}{2}}} ||\Pi_e^1 z_e||^2 / ||\Psi||^2_{-\frac{1}{2},\partial\Omega_e} \;; \\
\nu_e &= \inf_{V_e \in Q_k^1} ||\Pi_e^o \nabla V_e||_{\underset{\sim}{L}_2(\Omega_e)} / ||\nabla V_e||_{\underset{\sim}{L}_2(\Omega_e)} \\
\gamma_e &= \sup_{V_e \in Q_k^1} ||\nabla V_e - \Pi_e^o \nabla V_e||_{\underset{\sim}{L}_2(\Omega_e)} / ||\nabla V_e||_{\underset{\sim}{L}_2(\Omega_e)} \\
\beta(P) &= \min_{1 \le e \le E} \{\mu_e - \frac{\gamma_e}{2}, \nu_e - \frac{\gamma_e}{2}\}, \quad 0 \le \mu_e, \nu_e, \gamma_e \le 1
\end{aligned}
\tag{3.4}
$$

where $z_e \in H^1(\Omega_e)$ is the solution of an auxiliary problem, $-\nabla^2 z_e + z_e = 0$ in Ω_e and $\frac{\partial z_e}{\partial n} = \Psi$ on $\partial\Omega_e$, $\Psi \in Q_t^{-\frac{1}{2}}(\partial\Omega_e)$, and $||z_e||_{1,\Omega_e} = ||\Psi||^2_{-\frac{1}{2},\partial\Omega_e}$. We then have

Theorem 3.1. If $\beta(P) > 0$, there exists a unique finite element solution $\underset{\sim}{\Lambda}^o \in X_h$ of (3.4) and the following estimate holds:

$$
||e||_X \le Ch^\alpha ||u^o||_{H^\ell(P)}, \quad \ell \ge 2 \tag{3.5}
$$

where

$$
\underset{\sim}{e} = (u^o - U^o,\ \underset{\sim}{\nabla} u^o - \underset{\sim}{\Sigma}^o,\ -\frac{\partial u^o}{\partial n} - \Psi^o)
$$

$$
c = (1 + 2\sqrt{15}/\beta(P)) \max\{C_1, C_2, C_3\}, \quad \alpha = \min\{k, r+1, t+\frac{3}{2}, -1\}. \blacksquare
$$

We remark here that one easy way to guarantee that $\beta(P) > 0$ is to choose the polynomial spaces such that $k-1 \le r$ and $k \ge t+1$ (t even), $k \ge t+2$ (t odd) over a triangular element $\bar{\Omega}_e$.

A necessary condition can also be given.

Theorem 3.2. In order that (3.3) has a unique solution, it is necessary that $\mu_e > 0$, $1 \le e \le E$. $\blacksquare$

In the special case in which $Q_r^o(P) \subset \nabla(Q_k^1(P))$, then $\gamma_e = 0$, $\nu_e = 1$, and $\beta(P) = \min \mu_e = \mu$. Then $\mu > 0$ is a necessary and sufficient condition for the existence of a unique solution.

A necessary and sufficient condition that the parameters μ_e be positive is furnished by the rank condition.

THEOREM 3.3. The parameter μ_e in (3.4) is > 0 if and only if the following condition holds: for any $\Psi_e \in Q_t^{-\frac{1}{2}}(\Gamma)$,

$$\int_{\partial\Omega_e} \Psi_e U_e \, ds = 0 \quad \forall\, U_e \in Q_k^1(\Omega_e) \Rightarrow \Psi_e = 0 \tag{3.6}$$

■

Additional properties of such approximations are given in [4].

REFERENCES

1. Pian, T. H. H., "Element Stiffness Matrices for Boundary Compatibility and for Prescribed Boundary Stresses," Proceedings of the First Conference on Matrix Methods in Structural Mechanics, Wright-Patterson Air Force Base, 1965, AFFDL-TR-66-80, 1966, 457-477.

2. Brezzi, F., "Sur la Methode des Elements Finis Hybrides Pour le Probleme Biharmonique," Numer. Math. 24, 1975, 103-131.

3. Raviart, P. A., "Hybrid Finite Element Methods for Solving 2nd Order Elliptic Equations," Conference on Numerical Analysis, Royal Irish Academy, Dublin, 1974.

4. Babuska, I., Oden, J. T., and Lee, J. K., "Mixed Hybrid Finite Elements for Second-Order Elliptic Boundary-Value Problems," (to appear).

Author: J. T. Oden
Dept. of Aerospace Engr. and Engr. Mechanics
Engineering Lab Bldg. 305
The University of Texas at Austin
Austin, Texas 78712

Acknowledgement: The author's work on this project was supported through AFOSR Grant No. 74-2660.

ON SIMPLE INTERGRABILITY AND BOUNDED COARSE VARIATION

Charles F. Osgood and Oved Shisha

1. Let f be a complex function on $[0,\infty)$. Then [1,2] f is <u>simply integrable</u> iff there exists a number I with the following property: For every $\varepsilon > 0$ there are positive numbers B and Δ such that if $0 = x_0 < x_1 \ldots < x_n$, $x_n > B$; and $x_{k-1} \leq \xi_k \leq x_k$, $x_k - x_{k-1} < \Delta$ for $k = 1, 2, \ldots, n$, then

$$\left| I - \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) \right| < \varepsilon.$$

If f is simply integrable, then I is unique and equals $\int_0^\infty f(x)dx$, the improper Riemann integral of f on $[0,\infty)$. (By which we mean

$$\lim_{R\to\infty} \int_0^R f(x)dx \quad \text{(finite)}$$

where, for each $R > 0$, $\int_0^R f(x)dx$ is the proper Riemann integral of f on $[0,R]$.) If, on the other hand, f is improperly Riemann integrable on $[0,\infty)$, then it is simply integrable iff it is of <u>bounded coarse variation</u>. This means that for every $\varepsilon > 0$,

$$(1) \quad \sup \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| < \infty$$

The sup is taken over all sequences $(x_k)_{k=0}^{n}$, $n = 1, 2, \ldots$, with $0 \leq x_0 < x_1 < \ldots < x_n$; $x_k - x_{k-1} \geq \epsilon$, $k = 1,2,\ldots,n$.

The purpose of the present note is to further study the concepts of simple integrability and bounded coarse variation.

2. THEOREM 1. Let f be a complex function on $[0,\infty)$. The following statements are equivalent:

(i) For $k = 1,2,\ldots$, $\mathcal{O}_k = \sup_{k-1\le x<y\le k} |f(y) - f(x)| < \infty$, and $\Sigma_{k=1}^{\infty} \mathcal{O}_k < \infty$. (Here, "$\mathcal{O}$" stands for oscillation.)

(ii) f is of bounded coarse variation.

(iii) For some $\varepsilon > 0$, (1) holds.

Proof. Assume (i). Let $\varepsilon > 0$, $0 \le x_0 < x_1 < \ldots < x_n$, $n \ge 1$; $\min_{1\le k\le n} (x_k - x_{k-1}) \ge \varepsilon$. Consider the union of $\{x_0,x_1,\ldots,x_n\}$ and the set of integers N with $x_0 \le N \le x_n$, and arrange it as a strictly increasing sequence $(\xi_k)_{k=0}^{m}$. Then

$$\Sigma_{k=1}^{n}|f(x_k) - f(x_{k-1})| \le \Sigma_{k=1}^{m}|f(\xi_k) - f(\xi_{k-1})|$$

$$\le (2 + \varepsilon^{-1}) \Sigma_{k=1}^{\infty} \mathcal{O}_k.$$

Assume now (iii). Then f is clearly bounded in $[0,\infty)$. For $n, r = 1, 2, \ldots$ set

$$\mathcal{O}_{n,r} = \sup_{(n-1)r\le x<y\le nr} |f(y) - f(x)|.$$

If $r \ge 1$, then $\mathcal{O}_{nr+j} \le \mathcal{O}_{n+1,r}$ for $n = 0,1,2,\ldots$, and $j = 1,2,\ldots r$. Hence if for some $r \ge 1$ we show $\Sigma_{n=1}^{\infty}\mathcal{O}_{n,r} < \infty$, then it will clearly follow that $\Sigma_{k=1}^{\infty}\mathcal{O}_k < \infty$.

Let r be an integer $\ge 2\varepsilon$. We shall prove $\Sigma_{n=1}^{\infty}\mathcal{O}_{n,r} < \infty$. Suppose $\Sigma_{n=1}^{\infty}\mathcal{O}_{n,r} = \infty$. For $n = 1,2,\ldots$ choose x_n, y_n in $[(n-1)r, nr]$ such that $|f(x_n) - f(y_n)| \ge \mathcal{O}_{n,r}/2$, so that

$$|f(x_n) - f((n + \tfrac{1}{2})r)| + |f(y_n) - f((n + \tfrac{1}{2})r)| \geq O_{n,r}/2.$$

Then

$$\Sigma_{n=1}^{\infty} |f(x_n) - f((n + \tfrac{1}{2})r)| + |f(y_n) - f((n + \tfrac{1}{2})r)| = \infty.$$

Therefore at least one of the following diverges:

$$\Sigma_{n=1}^{\infty} |f(x_{2n}) - f((2n + \tfrac{1}{2})r)|,$$

$$\Sigma_{n=1}^{\infty} |f(y_{2n}) - f((2n + \tfrac{1}{2})r)|,$$

$$\Sigma_{n=1}^{\infty} |f(x_{2n+1}) - f((2n + \tfrac{3}{2})r)|,$$

$$\Sigma_{n=1}^{\infty} |f(y_{2n+1}) - f((2n + \tfrac{3}{2})r)|,$$

and hence we have at least one of the following:

$$\Sigma_{n=1}^{\infty} |f(x_{2n}) - f((2n - \tfrac{3}{2})r)| + |f((2n + \tfrac{1}{2})r) - f(x_{2n})| = \infty,$$

$$\Sigma_{n=1}^{\infty} |f(y_{2n}) - f((2n - \tfrac{3}{2})r)| + |f((2n + \tfrac{1}{2})r) - f(y_{2n})| = \infty,$$

$$\Sigma_{n=1}^{\infty} |f(x_{2n+1}) - f((2n - \tfrac{1}{2})r)|$$
$$+ |f((2n + \tfrac{3}{2})r) - f(x_{2n+1})| = \infty,$$

$$\Sigma_{n=1}^{\infty} |f(y_{2n+1}) - f((2n - \tfrac{1}{2})r)|$$
$$+ |f((2n + \tfrac{3}{2})r) - f(y_{2n+1})| = \infty.$$

However, each of these relations contradicts (iii).

3. For our second theorem we need the following

DEFINITION. *A sequence* $(x_n)_{n=0}^{\infty}$ *is called an allowable partition iff*

(2a) $0 = x_0 < x_1 < x_2 \ldots, x_n \to \infty$,

(2b) $0 < \inf_{k\geq 1}(x_k - x_{k-1}) \leq \sup_{k\geq 1}(x_k - x_{k-1}) < \infty$.

NOTATION. Given a complex function f on $[0,\infty)$, and given u,v, $0 \leq u < v$, we set $\mathcal{O}(f,u,v) = \sup_{u\leq x<y\leq v} |f(y) - f(x)|$; observe that if f is real on $[u,v]$, then

$$\mathcal{O}(f,u,v) = \sup_{u\leq x\leq v} f(x) - \inf_{u\leq x\leq v} f(x).$$

THEOREM 2. *Let* $(x_n)_{n=0}^{\infty}$ *be an allowable partition, and let* f *be a complex function, Riemann integrable on each* $[0,R]$, $0 < R < \infty$. *A necessary and sufficient condition for* f *to be simply integrable is that* $\Sigma_{n=1}^{\infty} f(\xi_n)(x_n - x_{n-1})$ *converges whenever* $x_{n-1} \leq \xi_n \leq x_n$, $n = 1,2,\ldots$.

To prove Theorem 2, we shall need the following

LEMMA. *Let* f *be a complex function on* $[0,\infty)$, *bounded on each* $[0,R]$, $0 < R < \infty$. *Suppose both* $(x_n)_{n=0}^{\infty}$ *and* $(y_n)_{n=0}^{\infty}$ *are allowable partitions, and one of the series* $\Sigma_{k=1}^{\infty}\mathcal{O}(f,x_{k-1},x_k)$, $\Sigma_{k=1}^{\infty}\mathcal{O}(f,x_{k-1},x_k)(x_k - x_{k-1})$ *converges. Then so do both* $\Sigma_{k=1}^{\infty}\mathcal{O}(f,y_{k-1},y_k)$, $\Sigma_{k=1}^{\infty}\mathcal{O}(f,y_{k-1},y_k)(y_k - y_{k-1})$.

4. *Proof of the Lemma*. Clearly $\Sigma_{k=1}^{\infty}\mathcal{O}(f,x_{k-1},x_k)$ converges, and it suffices to prove the convergence of $\Sigma_{k=1}^{\infty}\mathcal{O}(f,y_{k-1},y_k)$. Arrange the set $\{x_0,y_0,x_1,y_1,\ldots\}$ as a strictly increasing sequence $(z_k)_{k=0}^{\infty}$. Let $x_k = z_{m_k}$, $y_k = z_{n_k}$, $k = 0,1,2,\ldots$, and let M be a positive integer

such that each $[x_{k-1}, x_k]$ contains at most M z_n's. Then for $r = 1, 2, \ldots,$

$$\Sigma_{j=1}^{m_r} O(f, z_{j-1}, z_j) = \Sigma_{h=1}^{r} \Sigma_{j=m_{h-1}+1}^{m_h} O(f, z_{j-1}, z_j)$$

$$\leq (M - 1)\Sigma_{h=1}^{r} O(f, x_{h-1}, x_h) \leq (M - 1)\Sigma_{h=1}^{\infty} O(f, x_{h-1}, x_h).$$

Hence $\Sigma_{j=1}^{\infty} O(f, z_{j-1}, z_j)$ converges.

Let $k \geq 1$, $y_{k-1} \leq x < y \leq y_k$. Then, as is easily seen,

$$|f(y) - f(x)| \leq \Sigma_{j=n_{k-1}+1}^{n_k} O(f, z_{j-1}, z_j),$$

and hence, $O(f, y_{k-1}, y_k) \leq \Sigma_{n_{k-1}+1}^{n_k} O(f, z_{j-1}, z_j)$. Therefore for $N = 1, 2, \ldots,$

$$\Sigma_{k=1}^{N} O(f, y_{k-1}, y_k) \leq \Sigma_{k=1}^{N} \Sigma_{j=n_{k-1}+1}^{n_k} O(f, z_{j-1}, z_j)$$

$$\leq \Sigma_{j=1}^{\infty} O(f, z_{j-1}, z_j),$$

and hence, $\Sigma_{k=1}^{\infty} O(f, y_{k-1}, y_k) < \infty$.

5. Proof of Theorem 2. Necessity. Let $x_{n-1} \leq \xi_n \leq x_n$, $n = 1, 2, \ldots,$ and set

$$M_n = \int_{x_{n-1}}^{x_n} f(t)dt, \; n = 1, 2, \ldots.$$

As f is of bounded coarse variation, therefore, by Theorem 1, $\Sigma_{k=1}^{\infty} O_k < \infty$. Hence, by the Lemma, $\Sigma_{n=1}^{\infty} O(f, x_{n-1}, x_n)(x_n - x_{n-1})$ converges.

Let $\varepsilon > 0$. Let N be an integer ≥ 0 such that if $N \leq N_1 < N_2$, then

$$\left|\int_{x_{N_1}}^{x_{N_2}} f(t)dt\right| < \varepsilon/2$$

and

$$\Sigma_{n=N_1+1}^{N_2} O(f,x_{n-1},x_n)(x_n - x_{n-1}) < \varepsilon/2.$$

Let $N \leq N_1 < N_2$. Then

$$\left|\Sigma_{n=N_1+1}^{N_2} f(\xi_n)(x_n - x_{n-1})\right| \leq \left|\Sigma_{n=N_1+1}^{N_2} M_n\right|$$

$$+ \Sigma_{n=N_1+1}^{N_2} \left|f(\xi_n)(x_n - x_{n-1}) - M_n\right|$$

$$\leq \left|\int_{x_{N_1}}^{x_{N_2}} f(t)dt\right| + \Sigma_{n=N_1+1}^{N_2} O(f,x_{n-1},x_n)(x_n - x_{n-1}) < \varepsilon.$$

Hence $\Sigma_{n=1}^{\infty} f(\xi_n)(x_n - x_{n-1})$ converges.

Sufficiency. We may clearly assume that f is real on $[0,\infty)$. We first prove that f is improperly Riemann integrable on $[0,\infty)$.

Let n be a positive integer. If, for every x in $[x_{n-1},x_n]$, we had

$$f(x) > (x_n - x_{n-1})^{-1} \int_{x_{n-1}}^{x_n} f(t)dt,$$

then it would follow that

$$\int_{x_{n-1}}^{x_n} f(x)dx$$

is larger than itself [3, Theorem 388]. Let y_n satisfy $x_{n-1} \leq y_n \leq x_n$,

$$f(y_n) \leq \min[\{\inf_{x_{n-1}\leq x\leq x_n} f(x)\} + 2^{-n}, (x_n - x_{n-1})^{-1}\int_{x_{n-1}}^{x_n} f(t)dt].$$

Similarly, let $x_{n-1} \leq z_n \leq x_n$,

$$f(z_n) \geq \max\,[\{\sup_{x_{n-1} \leq x \leq x_n} f(x)\} - 2^{-n},\ (x_n - x_{n-1})^{-1} \int_{x_{n-1}}^{x_n} f(t)dt].$$

Then for $n = 1,2,\ldots,$

$$f(y_n)(x_n - x_{n-1}) \leq \int_{x_{n-1}}^{x_n} f(t)dt \leq f(z_n)(x_n - x_{n-1}),$$

and hence, if $0 \leq N_1 < N_2$, we have

$$\Sigma_{n=N_1+1}^{N_2} f(y_n)(x_n - x_{n-1}) \leq \int_{x_{N_1}}^{x_{N_2}} f(t)dt$$
$$\leq \Sigma_{n=N_1+1}^{N_2} f(z_n)(x_n - x_{n-1}).$$

Let $\varepsilon > 0$. Let N^* be an integer ≥ 0 such that if $N^* \leq N_1 < N_2$, then

$$\begin{aligned} &|\Sigma_{n=N_1+1}^{N_2} f(y_n)(x_n - x_{n-1})| < \varepsilon/9, \\ (3)\qquad &|\Sigma_{n=N_1+1}^{N_2} f(z_n)(x_n - x_{n-1})| < \varepsilon/9. \end{aligned}$$

Consequently if $N^* \leq N_1 < N_2$, then

$$|\int_{x_{N_1}}^{x_{N_2}} f(t)dt| < \varepsilon/9.$$

If $k > N^*$, then, setting $N_1 = k-1$, $N_2 = k$, we have, by (3),

$$|f(y_k)(x_k - x_{k-1})| < \varepsilon/9, \quad |f(z_k)(x_k - x_{k-1})| < \varepsilon/9.$$

Let $N^{**} \geq N^*$ be an integer such that if $k > N^{**}$, then $2^{-k}(x_k - x_{k-1}) < \varepsilon/9$.

If $k > N^{**}$, then throughout $[x_{k-1}, x_k]$,

$$f(y_k) - 2^{-k} \leq f(x) \leq f(z_k) + 2^{-k},$$

and hence,

$$|f(x)| \leq |f(y_k)| + |f(z_k)| + 2^{-k},$$

so that if $x_{k-1} \leq s \leq x_k$,

$$\left|\int_{x_{k-1}}^{s} f(x)dx\right| \leq (x_k - x_{k-1})[|f(y_k)| + |f(z_k)| + 2^{-k}]$$

$$< (\varepsilon/9) + (\varepsilon/9) + (\varepsilon/9) = \varepsilon/3,$$

and similarly,

$$\left|\int_{s}^{x_k} f(x)dx\right| < \varepsilon/3.$$

Let $x_{N^{**}} \leq s < t$, let x_{k_1} be the smallest $x_k > s$, and x_{k_2} the smallest $x_k > t$, so that $x_{k_1-1} \leq s < x_{k_1}$, $x_{k_2-1} \leq t < x_{k_2}$, $N^{**} < k_1 \leq k_2$. Then

$$\left|\int_s^t f(x)dx\right| \leq \left|\int_s^{x_{k_1-1}} f(x)dx\right| + \left|\int_{x_{k_1-1}}^{x_{k_2}} f(x)dx\right|$$

$$+ \left|\int_{x_{k_2}}^{t} f(x)dx\right| < (\varepsilon/3) + (\varepsilon/9) + (\varepsilon/3) < \varepsilon.$$

Hence f is improperly Riemann integrable on $[0,\infty)$.
For $k = 1,2,\ldots,$

$$\mathcal{O}(f, x_{k-1}, x_k) = \sup_{x_{k-1}\leq x\leq x_k} f(x) - \inf_{x_{k-1}\leq x\leq x_k} f(x) \leq f(z_k)$$

$$- f(y_k) + 2\cdot 2^{-k}.$$

Since $\Sigma_{k=1}^{\infty}(f(z_k) - f(y_k) + 2\cdot 2^{-k}) < \infty$, we have $\Sigma_{k=1}^{\infty}\mathcal{O}(f,x_{k-1},x_k) < \infty$. By the Lemma, $\Sigma_{k=1}^{\infty}\mathcal{O}(f,k-1,k) < \infty$, and hence, by Theorem 1, f is of bounded coarse variation. Therefore f is simply integrable.

6. We make here the following two observations:

(a) Let f be a complex function on $[0,\infty)$ for which $\{x \geq 0\colon f(x) \neq 0\}$ is unbounded. Then there exist sequences $(x_n)_{n=0}^{\infty}$: $0 = x_0 < x_1 < x_2 \ldots$, $x_n \to \infty$, and $(\xi_n)_{n=1}^{\infty}$: $x_{n-1} \leq \xi_n \leq x_n$, $n = 1,2,\ldots$, such that $\Sigma_{n=1}^{\infty}f(\xi_n)(x_n - x_{n-1})$ diverges.

(b) Let f be a complex function, improperly Riemann integrable on $[0,\infty)$. Then there exists a sequence $(x_n)_{n=0}^{\infty}$: $0 = x_0 < x_1 < x_2 \ldots$, $x_n \to \infty$, such that $\Sigma_{n=1}^{\infty}f(\xi_n)(x_n - x_{n-1})$ converges whenever $x_{n-1} \leq \xi_n \leq x_n$, $n = 1,2,\ldots$.

Let f^* be a simply integrable function for which $\{x \geq 0\colon f^*(x) \neq 0\}$ is unbounded, and let f^{**} be a complex function, improperly Riemann integrable on $[0,\infty)$, but not simply integrable. If we delete (2b) from the hypotheses of Theorem 2, then f^* provides a counterexample to the necessity part of Theorem 2, and f^{**} provides one to sufficiency.

7. The truth of (a) is readily seen. We indicate now the proof of (b). Let $f_1 = \mathrm{Re}f$, $f_2 = \mathrm{Im}f$. For $N = 0,1,2,\ldots$ choose

$$N = y_0^{(N)} < \ldots < y_{r_N}^{(N)} = N + 1$$

such that, for $j = 1,2$,

$$\Sigma_{k=1}^{r_n}\mathcal{O}(f_j,y_{k-1}^{(N)},y_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)}) < 2^{-N-1}.$$

If N is an integer ≥ 0, if $1 \leq p_N \leq q_N \leq r_N$, and if $y_{k-1}^{(N)} \leq \xi_k^{(N)} \leq y_k^{(N)}$ for $k = p_N$, p_N+1, $\ldots$, q_N, then

$$-2^{-N-1} < -\Sigma_{k=p_N}^{q_N} O(f_j, y_{k-1}^{(N)}, y_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)})$$

$$\leq \Sigma_{k=p_N}^{q_N} f_j(\xi_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)}) - \int_{y_{k-1}^{(N)}}^{y_k^{(N)}} f_j(x)dx$$

$$\leq \Sigma_{k=p_N}^{q_N} O(f_j, y_{k-1}^{(N)}, y_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)}) < 2^{-N-1} \quad (j = 1,2),$$

and hence,

$$|[\Sigma_{k=p_N}^{q_N} f_j(\xi_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)})] - \int_{y_{p_N-1}^{(N)}}^{y_{q_N}^{(N)}} f_j(x)dx| < 2^{-N-1}$$

$$(j = 1,2),$$

which implies

$$(4) \quad |[\Sigma_{k=p_N}^{q_N} f(\xi_k^{(N)})(y_k^{(N)} - y_{k-1}^{(N)})] - \int_{y_{p_N-1}^{(N)}}^{y_{q_N}^{(N)}} f(x)dx| < 2^{-N}.$$

Consider the sequence

$$0 = y_0^{(0)} <\ldots< y_{r_0}^{(0)} = y_0^{(1)} <\ldots< y_{r_1}^{(1)} = y_0^{(2)} < \ldots$$

and denote it $x_0, x_1, x_2, \ldots$ so that $0 = x_0 < x_1 < x_2 \ldots$, $x_n \to \infty$. Let $x_{n-1} \leq \xi_n \leq x_n$, $n = 1,2,\ldots.$ We show that $\Sigma_{n=1}^{\infty} f(\xi_n)(x_n - x_{n-1})$ converges.

Let $\varepsilon > 0$. Let $n_0 \geq 0$ be such that (i) if $n \geq n_0$, then $x_n \geq N_0$, where N_o is an integer ≥ 0 with $2^{-N_0} < \varepsilon/4$; and (ii)

$$|\int_{x_s}^{x_t} f(x)dx| < \varepsilon/2$$

whenever $n_o \leq s < t$. Let $n_o \leq s < t$; we claim:

$$|\Sigma_{n=s+1}^{t} f(\xi_n)(x_n - x_{n-1})| < \varepsilon.$$

Indeed, set

$$S = [\Sigma_{n=s+1}^{t} f(\xi_n)(x_n - x_{n-1})] - \int_{x_s}^{x_t} f(x)dx.$$

Then

$$|\Sigma_{n=s+1}^{t} f(\xi_n)(x_n - x_{n-1})| \leq |S| + \left|\int_{x_s}^{x_t} f(x)dx\right|$$

$$< |S| + (\varepsilon/2).$$

Denote $N_1 = [x_s]$, so that $2^{-N_1} < \varepsilon/4$, and $[x_{t-1}] = N_1 + u$, $u \geq 0$. Then, using (4), one can show:

$$|S| < \Sigma_{k=0}^{u} 2^{-N_1-k} < 2\cdot 2^{-N_1} < \varepsilon/2.$$

Hence

$$|\Sigma_{n=s+1}^{t} f(\xi_n)(x_n - x_{n-1})| < \varepsilon.$$

References

1 Haber, S. and O. Shisha, An integral related to numerical integration, Bull. Amer. Math. Soc. 79 (1973), 930-932.

2 Haber, S. and O. Shisha, Improper integrals, simple integrals, and numerical quadrature, J. Approximation Theory 11 (1974), 1-15.

3 Landau, E., Differential and Integral Calculus, Chelsea, New York, 2nd edition, 1960.

Charles F. Osgood
Naval Research Laboratory
Washington, D.C. 20375

Oved Shisha
Department of Mathematics
University of Rhode Island
Kingston, R.I. 02881

SHAPE PRESERVING SPLINE INTERPOLATION

Eli Passow and John A. Roulier

For a set of increasing (and/or convex) data, we can consider the possibility of finding a spline interpolant, of preassigned smoothness, which is increasing (and/or convex). To carry out this investigation we construct an auxiliary set of points and use the monotonicity and convexity preserving properties of Bernstein polynomials.

1 Introduction

Let $\Delta = \{x_0 < x_1 < \ldots < x_N\}$, let $j \leq n$, and let $S_n^j = S_n^j(\Delta)$ be the set of splines of degree n and deficiency n-j, with knots at x_i, $i = 1,2,\ldots,N-1$. That is, $f \in S_n^j(\Delta)$ if $f \in C^j[x_0,x_N]$ and on each of the intervals $[x_{i-1},x_i]$, $i = 1, 2,\ldots,N$, $f \in P_n$, the set of algebraic polynomials of degree $\leq n$.

Let $y_0 < y_1 < \ldots < y_N$. For fixed j and n we ask: Does there exist an increasing $f \in S_n^j(\Delta)$ satisfying $f(x_i) = y_i$, $i = 0,1,\ldots,N$? More precisely, what conditions on the data (x_i,y_i) guarantee the existence of such an interpolant? Similarly, let $s_i = (y_i-y_{i-1})/(x_i-x_{i-1})$, $i = 1,2,\ldots,N$, and suppose that $s_{i-1} < s_i$, $i = 2,3,\ldots,N$. In this case where we say the data are convex, what additional conditions guarantee the existence of a convex interpolant in $S_n^j(\Delta)$?

We can approach this problem from a slightly different point of view. For fixed monotone or convex data and fixed j we may ask: How large must n be to guarantee the existence of a monotone or convex interpolant in $S_n^j(\Delta)$? (For $j = n$; i.e., polynomials of degree n, an answer to this question in

Preceding page blank

the monotone case was given in [4,5].) The advantage of this approach, which we adopt, is that by minimizing n we economize on the number of parameters needed to obtain a shape preserving interpolant of desired smoothness (class C^j).

2 The main results

We will state our results, with some indication of proofs. The following definition will be important in our investigation.

DEFINITION 1. Suppose the data (x_i, y_i), $i = 0,1,\ldots,N$, are non-decreasing (and/or non-concave). Let $\bar{x}_i = x_i + \alpha_i \Delta x_i$, where $0 < \alpha_i < 1$ and $\Delta x_i = x_i - x_{i-1}$, $i = 1,2,\ldots,N$. The numbers t_i, $i = 1,2,\ldots,N$, are said to be *increasing (and/or convex) $\{\alpha_i\}$-admissible* for the data (x_i, y_i) if the piecewise linear function $L(x)$ generated by the points (x_0, y_0), $(\bar{x}_1, t_1)$, $(\bar{x}_2, t_2), \ldots, (\bar{x}_N, t_N)$, (x_N, y_N), passes through the points (x_i, y_i), $i = 1,2,\ldots,N-1$, and is non-decreasing (and/or non-concave). If $\alpha_i = \alpha$ for all i, then we say that the numbers t_i are *α-admissible*.

LEMMA 1. *Let* k,m,n *be positive integers, with* $m < n$. *Let* $p \in P_{kn}$ *and suppose that* $p^{(j)}(a) = 0$, $j = 2,3,\ldots,mk$, *and* $p^{(j)}(b) = 0$, $j = 2,3,\ldots,(n-m)k$, *where* $a < b$. (*If* $mk = 1$ *or* $(n-m)k = 1$, *then the corresponding condition is vacuous*.) *Then the tangent lines to the graph of* p *at* $x = a$ *and* $x = b$ *intersect at* $\bar{x} = a + m(b-a)/n$. *In particular, if* $n = 2m$, *then* $\bar{x} = (a + b)/2$.

For the sake of simplicity, we state our first theorem for $\alpha_i = \alpha$, but a similar result for general α_i holds.

THEOREM 1. *Let* m,n *be natural numbers, with* $n \geq 2m$, *and let* $\alpha = m/n$. *Then there exist increasing (and/or convex) α-admissible numbers for the data* (x_i, y_i) *if and only if for all* $k \geq 1$

there exists $f \in S_{kn}^{km}(\Delta)$ *satisfying*

$$f(x_i) = y_i,\ i = 0,1,\ldots,N;$$

$$f_+^{(j)}(x_i) = 0,\ j = 2,3,\ldots,mk;\ i = 1,2,\ldots,N-1;$$

$$f_-^{(j)}(x_i) = 0,\ j = 2,3,\ldots,\ (n-m)k;\ i = 1,2,\ldots,N-1;$$

$$f'(x) \geq 0,\ x \in [x_0,x_N]\ (\text{and/or } f''(x) \geq 0,\ x \in [x_0,x_N]).$$

Thus, in particular, $f^{(j)}(x_i) = 0$, $j = 2,3,\ldots,mk$; $i = 1,2,\ldots,N-1$.

Proof. Suppose $\{t_i\}$ are α-admissible, and let $L(x)$ be the associated piecewise linear function in Definition 1. Let $q_i(x)$ be the Bernstein polynomial of degree kn of the restriction of $L(x)$ to the interval $[x_{i-1},x_i]$, $i = 1,2,\ldots,N$, and let $f(x) = q_i(x)$, $x \in [x_{i-1},x_i]$, $i = 1,2,\ldots,N$. It follows from properties of the Bernstein polynomials [1, pp. 13, 23] that $f \in S_{kn}^{km}(\Delta)$, $f(x_i) = y_i$, $i = 0,1,\ldots,N$, and f is increasing (and/or convex).

The converse follows easily from Lemma 1.

The existence of an α-admissible set is thus essential to the solution of our problem. The next result is an algorithm which is useful in finding such sets.

α-ALGORITHM. Let $m_1 = s_1$, $M_1 = \min\{s_2, s_1/(1-\alpha)\}$, and for $i = 2,3,\ldots,N-1$ let $m_i = (s_i-\alpha M_{i-1})/(1-\alpha)$, $M_i = \min\{s_{i+1}, (s_i-\alpha m_{i-1})/(1-\alpha)\}$.

THEOREM 2. *For non-decreasing, non-concave data* (x_i,y_i), $i = 0,1,\ldots,N$, *there exist increasing and convex* α-*admissible numbers* $\{t_i\}$ *if and only if the* α-*algorithm can be carried out with* $m_i \leq s_{i+1}$, *for* $i = 2,3,\ldots,N-1$.

As a corollary to Theorem 2 we have the following, which is a generalization of a result for quadratic splines in [3].

THEOREM 3. *If* $0 < s_1 < \ldots < s_N$ $(s_{i+1}-2s_i+s_{i-1} \geq 0,\ i = 2, 3,\ldots,N-1)$, *then there exists a set of increasing (convex) $\frac{1}{2}$-admissible points for* (x_i,y_i). *Hence, in this case, for all* n, *there exists an increasing (convex) interpolant in* $S^n_{2n}(\Delta)$.

If we restrict our attention to monotone interpolation, it has been shown [2] that for fixed j we may always choose $n = 2j+1$. For convex interpolation, however, no such result is possible, because of the following.

THEOREM 4. *For any integer* $n \geq 1$, *there exists a convex set of data points* (x_i,y_i), $0 \leq i \leq 4$, *for which no* $f \in S^1_n(\Delta)$ *satisfies* $f(x_i) = y_i$, $0 \leq i \leq 4$, *with* f *convex on* $[x_0,x_4]$.

The next theorem extends the result of [2] to S^n_{2n}, but additional knots must be allowed. The proof is an application of Theorem 1.

THEOREM 5. *Let* $y_{i-1} < y_i$, $\bar{x}_i = (x_{i-1}+x_i)/2$, $\bar{y}_i = (y_{i-1}+y_i)/2$, $i = 1,2,\ldots,N$, *and let* $\Delta = \{x_0,\bar{x}_1,x_1,\bar{x}_2,\ldots,\bar{x}_N,x_N\}$. *Then, for all* $n \geq 1$, *there exists an increasing* $f \in S^n_{2n}(\Delta)$ *such that* $f(x_i) = y_i$, $i = 0,1,\ldots,N$, *and* $f(\bar{x}_i) = \bar{y}_i$, $i = 1,2,\ldots,N$.

Theorem 5 holds for piecewise monotone interpolation as well (see [2]).

3 Conclusion

In closing we point out that since Theorem 1 yields no better than $f \in S^n_{2n}$, any attempt to decrease the deficiency of the interpolating spline will require a new approach.

References

[1] Lorentz, G.G., *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953.

[2] Passow, E., Piecewise monotone spline interpolation, J. Approximation Theory *12* (1974), 240-241.

[3] Passow, E., Monotone quadratic spline interpolation, J. Approximation Theory (to appear).

[4] Passow, E., An improved estimate of the degree of monotone interpolation, J. Approximation Theory (to appear).

[5] Passow, E. and L. Raymon, The degree of piecewise monotone interpolation, Proc. Amer. Math. Soc. 48 (1975), 409-412.

Eli Passow*
Department of Mathematics
Temple University
Philadelphia, Pa. 19122

John A. Roulier
Department of Mathematics
North Carolina State University
Raleigh, N.C. 27607

* Supported in part by a Temple University Grant-In-Aid of Research.

NEGATIVE THEOREMS ON CONVEX APPROXIMATION

Eli Passow and John A. Roulier

In this paper we show that there exist functions $f \in C[-1, +1]$ with all r-th order divided differences uniformly bounded away from zero for r fixed for which infinitely many of the polynomials of best approximation to f do not have nonnegative r-th derivatives on $[-1, +1]$.

1 Introduction

In [1] - [5] there appear many examples of functions f in $C[a,b]$ with nonnegative r-th divided differences there for which infinitely many of the polynomials of best approximation to f fail to have nonnegative r-th derivatives. None of these examples has the r-th divided differences uniformly bounded away from zero. In [7] Roulier shows that if $f \in C^{2r}[-1, +1]$ and if $f^{(r)}(x) \geq \delta > 0$ on $[-1, 1]$ then for n sufficiently large the polynomial of best approximation of degree less than or equal to n has a positive r-th derivative on $[-1, +1]$.

On the other hand for the case $r = 1$ Roulier in [6] shows that first divided differences of f uniformly bounded away from zero is not sufficient to insure that for n sufficiently large the polynomial of best approximation to f is increasing on $[-1, 1]$.

In this paper we extend the results of [6] to the case when $r \geq 1$. The proofs are similar to those in [6] but make use of higher order divided differences and their properties. They will appear elsewhere.

Preceding page blank

2 Main Theorems

The following theorems treat the situations where $f[x_0,\ldots,x_{r+1}] \geq \delta > 0$ for $-1 \leq x_0 <\ldots< x_{r+1} \leq 1$ and either $f \in C^{r-1}[-1, +1]$ or $f \in C^r[-1, +1]$. The cases $f \in C^j[-1, +1]$ $j = r + 1,\ldots,2r - 1$ are unsolved. As mentioned above, the case $f \in C^{2r}[-1, +1]$ has been solved [7]. In what follows p_n will denote the polynomial of best approximation to f on $[-1, +1]$.

THEOREM 1. Let $f \in C[-1, 1]$ have bounded r-th order divided differences (if $f \in C^r[-1, +1]$, then this happens) and non-negative $r + 1^{st}$ order divided differences on $[-1, +1]$. Assume that there is no $C > 0$ for which

$$E_n(f) \leq C/(n + 1)^{r+1} \quad \text{for} \quad n = 0,1,\ldots \quad .$$

Then there are infinitely many n for which we do not have $p_n^{(r+1)}(x) \geq 0$ on $[-1, +1]$.

COROLLARY 2(a). Let $f \in C^r[-1, +1]$ and assume that f has nonnegative $r + 1^{st}$ order divided differences on $[-1, +1]$. Define $g(t) = f(\cos t)$. Assume that

$$\text{(1)} \quad \limsup_{k \to \infty} k^{r+1} \omega_{r+1}(g, \tfrac{1}{k})/\log k = +\infty.$$

Then there are infinitely many n for which we do not have $p_n^{(r+1)}(x) \geq 0$ on $[-1, +1]$.

COROLLARY 2(b). If f has nonnegative $(r + 1)$-st order divided differences on $(-1 - \varepsilon, 1 + \varepsilon)$ for some $\varepsilon > 0$ and if there is no $C > 0$ for which

$$E_n(f) \leq C/(n + 1)^{r+1} \quad \text{for} \quad n = 0, 1,\ldots$$

then there are infinitely many n for which we do not have

$$p_n^{(r+1)}(x) \geq 0 \quad \text{on} \quad [-1, +1].$$

THEOREM 3. *Let* $f \in C^{r-1}[-1, +1]$ *and assume that* f *has nonnegative* $r + 1^{st}$ *order divided differences. Assume that there is no* $C > 0$ *for which*

$$E_n(f) \leq C/(n + 1)^r \quad \text{for} \quad n = 0, 1, \ldots \ .$$

Then there are infinitely many n *for which we do not have* $p_n^{(r+1)}(x) \geq 0$ *on* $[-1, +1]$.

COROLLARY 4. *Let* $f \in C^{r-1}[-1, +1]$ *and assume that* f *has nonnegative* $r + 1^{st}$ *order divided differences. Define*

$$g(t) = f(\cos t).$$

Assume that

$$\limsup_{k \to \infty} k^r \omega_r(g, \tfrac{1}{k})/\log k = +\infty. \tag{2}$$

Then there are infinitely many n *for which we do not have* $p_n^{(r+1)}(x) \geq 0$ *on* $[-1, +1]$.

THEOREM 5. *Given integer* $r \geq 0$ *and modulus of continuity* ω *there exists* $f \in C^r[-1, +1]$ *with*

$$f[x_0,\ldots,x_{r+1}] \geq \delta > 0 \quad \text{for all} \quad x_0 < \ldots < x_{r+1} \tag{3}$$

in $[-1, +1]$ *and with*

$$\omega(h) \leq \omega(f^{(r)}, h) \leq K\omega(h) \tag{4}$$

and yet there are infinitely many n *for which we do not have* $p_n^{(r+1)}(x) \geq 0$.

THEOREM 6. *Given integer* $r \geq 1$ *and modulus of continuity* ω *there exists* $f \in C^{r-1}[-1, +1]$ *with*

$$f[x_0,\ldots,x_{r+1}] \geq \delta > 0 \quad \text{for all} \quad x_0 < \ldots < x_{r+1} \tag{5}$$

in $[-1, +1]$ *and with*

$$\omega(h) \leq \omega(f^{(r-1)}, h) \leq K\omega(h)$$

<u>and yet there are infinitely many</u> n <u>for which we do not have</u> $p_n^{(r+1)}(x) \geq 0$.

References

1 Lorentz, G. G., Monotone Approximation, "Inequalities III", (O. Shisha, Ed.), Academic Press Inc., (1969), 201-215.

2 Lorentz, G. G. and K. L. Zeller, Monotone approximation by algebraic polynomials, Trans. Amer. Math. Soc. 149 (1970) 1-18.

3 Lorentz, G. G. and K. L. Zeller, Degree of approximation by monotone polynomials II, J. Approximation Theory 2 (1969) 265-269.

4 Roulier, J. A., Monotone and weighted approximation, Dissertation, Syracuse University, 1968.

5 Roulier, J. A., Polynomials of best approximation which are monotone, J. Approximation Theory 9 (1973), 212-217.

6 Roulier, J. A., Negative theorems on monotone approximation, Proc. Amer. Math. Soc., (to appear).

7 Roulier, J. A., Best approximation to functions with restricted derivatives, J. Approximation Theory, (to appear).

Eli Passow *
Department of Mathematics
Temple University
Philadelphia, Pennsylvania 19122

John A. Roulier
Department of Mathematics
North Carolina State University
Raleigh, North Carolina 27607

*Supported in part by a Temple University Grant-In-Aid of Research.

APPROXIMATION BY CIRCLES

T. J. Rivlin

The problem considered is to assign a measure of circularity to a given compact set in the plane. The measure adopted is the size of the smallest annulus containing the given set. Two different notions of the size of an annulus, that of area and that of difference of radii, are studied.

1. Introduction

The problem we consider is to assign a measure of circularity to a given compact set in the plane. Our approach is to determine the best annulus which contains the given set and rate the circularity of the given set according to the size of the annulus. Of course, we must make precise the sense of "best" and "size" in the preceding sentence. We consider two different assignments of size to an annulus. The first is its area, the second the difference of its radii. A notion of size having been fixed we define a best annular approximation to a given compact set in the plane, S, to be an annulus of least size which contains S.

It turns out that with the criterion of size corresponding to area the problem of best annular approximation of S is equivalent to a well-known linear uniform approximation problem. This is discussed in Section 2. The second notion of size is equivalent to a somewhat novel non-linear approximation problem as we shall see in Section 3.

What follows is a summary of selected results in this problem area. A full report including all details will be published elsewhere. The author wishes to thank his colleagues A. J. Hoffman and C. A. Micchelli for valuable suggestions.

2. Least area criterion

Given a compact set S in the plane, for any complex number, w, put

(1) $r_2(w) = \max_{z\in S}|z-w|$; $r_1(w) = \min_{z\in S}|z-w|$.

Clearly S is contained in the annulus

(2) $r_1(w) \le |z-w| \le r_2(w)$

whose area is $A(w) = \pi(r_2{}^2-r_1{}^2)$.

If

$$\inf_w A(w) = A(w_o)$$

then the annulus described by (2) with $w = w_o$ is a best annulus in the sense of area which contains S.

Now put $f(x,y) = x^2+y^2$ and let V be the linear space of linear functions $ax+by+c$. $f(x,y)$ has a best uniform approximation out of V on S.

THEOREM 1. If $w_o:(x_o,y_o)$ is the center of a best annulus in the sense of area for S, then

$$v_o(x,y) = 2x_ox+2y_oy-(x_o{}^2+y_o{}^2-\frac{r_1{}^2(w_o)+r_2{}^2(w_o)}{2})$$

is a best approximation to $f(x,y)$ on S out of V, and $M = ||f-v_o|| = (r_2{}^2(w_o)-r_1{}^2(w_o))/2$. Conversely, if $v_o(x,y) = 2x_ox+2y_oy+c_o$ is a best uniform approximation to $f(x,y)$ on S out of V then w_o: (x_o,y_o) is the center of a best annulus for S, and if $M = ||f-v_o||$, then $r_2{}^2(w_o) = c_o+x_o{}^2+y_o{}^2+M$; $r_1{}^2(w_o) = c_o+x_o{}^2+y_o{}^2-M$, and hence $r_2{}^2(w_o)-r_1{}^2(w_o) = 2M$.

This equivalence establishes the existence of a best annular approximation and sheds light on uniqueness questions. The characterization of best linear approximation here is

elegant and simple in terms of the notion of extremal signature (cf. Rivlin and Shapiro, SIAM Journal, 9, 1961, 670-699). It enables us, for example, to determine best annular approximations to sets S consisting of a finite number of points, and to establish the following geometric fact. If S is an oval then there is a pair of concentric circles of which the inner one lies inside the oval and touches it at at least two distinct points, and the outer one lies outside the oval and touches it at at least two distinct points. The question of what is the most general class of simple closed curves having this property is open and seems difficult.

3. Difference of radii criterion

In this section we retain the notation of Section 2, but here instead of $A(w)$ we are interested in $r(w) = r_2(w)-r_1(w)$. If

$$\inf_w r(w) = r(w_o)$$

we call the annulus described by (2) with $w = w_o$ a best approximation in the uniform sense to S.

We note at once that a best uniform annular approximation need not exist. For example, it does not exist if S consists of 3 or more collinear points. However, a uniform approximation problem equivalent to the problem of best uniform annular approximation does exist.

Consider

$$(3)\quad F(x,y;\, h,k,t) = ((x-h)^2+(y-k)^2)^{1/2} - t.$$

Then the following equivalence is quite obvious.

LEMMA 1.

$$\mu = \min_{h,k,t}\ \max_{(x,y)\in S} |F(x,y;h,k,t)| = \max_{(x,y)\in S} |F(x,y;\, \bar{h},\bar{k},\bar{t})|$$

if, and only if, the annulus centered at $\bar{w}:(\bar{h},\bar{k})$ *with*

$r_1(\bar{w}) = \bar{t}-\mu$ and $r_2(\bar{w}) = \bar{t}+\mu$ is a best uniform annular approximation to S.

Thus the equivalent non-linear approximation problem referred to is that of finding a best approximation on S to zero by means of functions of the form (3). A familiar kind of argument now yields the following necessary condition.

THEOREM 2. If $\bar{w}$: $(\bar{h},\bar{k})$ is the center of an annulus which is a best uniform approximation to S, with $0 < r_1(\bar{w}) < r_2(\bar{w})$, then there exist a positive integer $k \le 4$, positive numbers $\lambda_1,\dots,\lambda_k$, points $z_1,\dots,z_k$ of S and disjoint subsets I_1 and I_2 of $\{1,\dots,k\}$ such that $|I_2|+|I_1| = k$,

$$\sum_{j=1}^{k} \lambda_j = 1,$$

$$\sum_{j\in I_1} \lambda_j = \sum_{j\in I_2} \lambda_j,$$

and

$$\sum_{j\in I_1} \lambda_j \frac{z_j-\bar{w}}{r_1(\bar{w})} = \sum_{j\in I_2} \lambda_j \frac{z_j-\bar{w}}{r_2(\bar{w})} ,$$

where $|z_j-\bar{w}| = r_1$ for $j\in I_1$ and $|z_j-\bar{w}| = r_2$ for $j\in I_2$.

In certain simple examples these conditons give a good deal of information about best approximations. For instance, when S consists of 4 points a best uniform annulus for S has 2 points of S on its outer boundary and 2 on its inner boundary; thus, in view of the results mentioned in Section 2, it cannot coincide with a best annulus in the sense of area in many cases.

Sufficient conditions for best uniform annular approximation seem hard to come by and are under investigation.

T. J. Rivlin
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

THE BEHAVIOR OF THE PADÉ TABLE FOR THE EXPONENTIAL

E.B. Saff and R.S. Varga

In this paper we survey recent results and present some new theorems on the behavior of Padé approximants for e^{-z}. The new results include necessary and sufficient conditions for (1) a sequence of approximants to be pole-free in an infinite sector, and (2) a sequence of approximants to converge geometrically in the uniform norm over an infinite sector.

1 Introduction

While the study of the Padé table for the exponential function dates back to Padé's thesis, there has been renewed interest in the subject because of its usefulness in certain numerical schemes for solving parabolic differential equations. Several recent papers have appeared which consider the questions of location of zeros and poles, regions of convergence, and degree of convergence of sequences from the table (see [3], [9], [10], [14], [16]). The purpose of the present paper is to survey some of these results and also to establish some new theorems. In this first section we introduce the necessary notation, in Sec. 2 we discuss zero and pole-free regions, and in Sec. 3 we consider the degree of convergence of Padé approximants in unbounded regions.

To be specific we shall deal with the complex negative exponential function e^{-z}. For each pair (ν,n) of nonnegative integers the Padé approximant $R_{\nu,n}(z)$ of type (ν,n) for e^{-z} is defined as that unique rational function with numerator degree ν, denominator degree n, which has greatest contact with e^{-z} at the origin, i.e.,

$$(1.1)\quad e^{-z} - R_{\nu,n}(z) = \mathcal{O}(z^{n+\nu+1}) \quad \text{as} \quad z \to 0 \ .$$

Explicitly it is known [6] that $R_{\nu,n}(z) = Q_{\nu,n}(z)/P_{\nu,n}(z)$, where

$$(1.2)\quad Q_{\nu,n}(z) = \sum_{j=o}^{\nu} \frac{(n+\nu-j)!\nu!(-z)^j}{(n+\nu)!j!(\nu-j)!} \ ,$$

and

$$(1.3)\quad P_{\nu,n}(z) = \sum_{j=o}^{n} \frac{(n+\nu-j)!n!z^j}{(n+\nu)!j!(n-j)!} \ .$$

The polynomials $Q_{\nu,n}(z)$ and $P_{\nu,n}(z)$ are referred to respectively as the <u>Padé numerator</u> and <u>Padé denominator</u> of type (ν,n) for e^{-z} . From the representations (1.2) and (1.3) it is apparent that $Q_{\nu,n}(z) = P_{n,\nu}(-z)$, and so any result on the location of the poles of Padé approximants to e^{-z} has a reformulation in terms of zeros.

The Padé approximants $R_{\nu,n}(z)$ are usually studied in the context of the following doubly infinite array known as the <u>Padé table</u>:

$$(1.4)\quad \begin{bmatrix} R_{0,0} & R_{1,0} & R_{2,0} & \cdots \\ R_{0,1} & R_{1,1} & R_{2,1} & \cdots \\ R_{0,2} & R_{1,2} & R_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} .$$

Notice that the first row of the table consists of the partial sums $R_{\nu,o}(z) = \sum_{k=o}^{\nu} (-z)^k/k!$ of e^{-z} and that the first column is composed of the reciprocals of the partial sums for the positive exponential, i.e., $R_{o,n}(z) = [\sum_{k=o}^{n} z^k/k!]^{-1}$.

2 Unbounded pole-free regions

The asymptotic behavior of the poles of the first column of (1.4), i.e., the zeros of the partial sums $s_n(z)=\sum_{k=o}^{n} z^k/k!$, was studied by Szegö [13] and Dieudonné [2]. As a consequence of their results it follows that any infinite sector with vertex at the origin contains infinitely many poles of the sequence $\{R_{o,n}(z)\}_{n=o}^{\infty}$. By way of contrast it is shown in R.S. Varga's thesis [17] that the infinite half-strip $|y|\leq\sqrt{6}$, $x\geq 0$, is free of poles of the whole sequence $\{R_{o,n}(z)\}_{n=o}^{\infty}$. More recently, Newman and Rivlin ([4], [5]) established that there exists an unbounded parabolic region, namely

$$(2.1) \quad y^2 \leq dx \quad , \quad x \geq 0 \quad , \ d \doteq 0.745 \quad ,$$

which is pole-free for the sequence $\{R_{o,n}(z)\}_{n=o}^{\infty}$. Furthermore they proved that for this sequence parabolic growth characterizes the largest pole-free region symmetric about the positive real axis.

Using continued fraction techniques the authors were able to improve upon the result of (2.1) and also to obtain similar results for all the columns of the table (1.4). In stating this theorem it is convenient to introduce the <u>normalized Padé approximants</u> $R_{\nu,n}((\nu+1)z)$.

THEOREM 2.1. (Saff, Varga [8], [11]). <u>For all $\nu\geq 0$, $n\geq 0$, the normalized Padé approximant $R_{\nu,n}((\nu+1)z)$ has no poles in the unbounded parabolic region</u>

$$(2.2) \quad P_1:=\{z = x + iy : y^2 \leq 4(x + 1) \quad , x > -1\} \quad .$$

<u>Moreover, every boundary point of P_1 is a limit point of poles of the collection $\{R_{\nu,n}((\nu+1)z)\}_{\nu=o,n=o}^{\infty\ ,\infty}$</u>.

In particular, Theorem 2.1 implies that the first column

of the table (1.4), for which $\nu=0$, is pole-free in P_1 (a region larger than that of (2.1)) and, in general, the $(\nu+1)$st column $\{R_{\nu,n}(z)\}_{n=0}^{\infty}$ is pole-free in the parabolic region

$$(2.3)\quad P_{\nu+1} := \{z=x+iy : y^2 \leq 4(\nu+1)(x+\nu+1),\ x > -(\nu+1)\} .$$

These facts have proved useful in approximation estimates for the matrix exponential as discussed in a recent paper of Van Loan [15].

While Theorem 2.1 is sharp, it does not include the fact that for an arbitrary fixed ν the largest pole-free region for the sequence $\{R_{\nu,n}(z)\}_{n=0}^{\infty}$ has parabolic growth. We shall prove this in

THEOREM 2.2. *For each fixed* $\nu \geq 0$, *the Padé approximant* $R_{\nu,n}(z)$ *for* e^{-z} *has a pole of the form*

$$(2.4)\quad (n+\sqrt{n}\,x_{\nu,n}) + i\sqrt{n}\,y_{\nu,n},\ \text{where}\ x_{\nu,n} + iy_{\nu,n} \to w_\nu (\neq 0)$$

as $n\to\infty$.

Note that as

$$\lim_{n\to\infty} \frac{ny_{\nu,n}^2}{n+\sqrt{n}\,x_{\nu,n}} = (\operatorname{Im} w_\nu)^2 ,$$

there are poles of the $R_{\nu,n}(z)$ which asymptotically fall on the parabolic arc $y^2=(\operatorname{Im} w_\nu)^2 x$, as $n\to\infty$. When $\nu=0$, Theorem 2.2 reduces to the known result of Newman and Rivlin [4]. The proof of Theorem 2.2 requires the following lemma:

LEMMA 2.1. *For each nonnegative integer* ν, *the function*

$$(2.5)\quad F_\nu(z) := \int_0^\infty t^\nu e^{-zt-t^2/2}dt ,\quad (0\leq t<\infty) ,$$

is an entire function having at least one (finite) zero $w_\nu (\neq 0)$.

Proof. It is easy to see that $F_\nu(z)$ is entire. More precisely, on writing

$$(2.6)\quad F_\nu(z) = \sum_{k=o}^{\infty} a_k(\nu)z^k \quad ,$$

it follows from (2.5) that

$$(2.7)\quad a_k(\nu) = \frac{(-1)^k\, 2^{(\frac{k+\nu-1}{2})}\,\Gamma(\frac{k+\nu+1}{2})}{k!} \quad , \quad \underline{k\geq 0} \quad .$$

Using Stirling's formula one can verify from (2.7) that $F_\nu(z)$ is of order 2 for each ν and, moreover, $F_\nu(z)$ is an entire function of <u>perfectly</u> <u>regular</u> <u>growth</u>; specifically if $M_{F_\nu}(r) := \max_{|z|=r} |F_\nu(z)|$, then

$$(2.8)\quad \lim_{r\to\infty} \frac{\ln M_{F_\nu}(r)}{r^2} = \frac{1}{2} \quad .$$

Now assume to the contrary that $F_\nu(z)$ has no zeros. By the Hadamard Factorization Theorem ([1,p.22]), we can express $F_\nu(z)$ as $F_\nu(z)=e^{q(z)}$, where $q(z)$ is a polynomial of degree not exceeding the order of F_ν . Hence, since F_ν is of order 2 , there exist constants α_1,α_2 such that

$$(2.9)\quad F_\nu(z) = F_\nu(0)e^{\alpha_1 z+\alpha_2 z^2} = a_o(\nu)e^{\alpha_1 z+\alpha_2 z^2} \quad , \text{ for all } z \quad .$$

Using (2.7), (2.8), and the fact that $M_{F_\nu}(r) = F_\nu(-r)$, it is easy to show that

$$\alpha_2 = \frac{1}{2} \quad , \quad \alpha_1 = -\sqrt{2}\,\Gamma(\tfrac{\nu+2}{2})/\Gamma(\tfrac{\nu+1}{2}) \quad ,$$

and hence the right-hand member of (2.9) is completely specified. Equating the coefficients of z^2 in (2.9) results in the equation

$$\frac{2^{\frac{\nu+1}{2}}\,\Gamma(\frac{\nu+3}{2})}{2} = \frac{2^{\frac{\nu-1}{2}}\,\Gamma(\frac{\nu+1}{2})}{2}\left\{1 + \left[\frac{\sqrt{2}\,\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+1}{2})}\right]^2\right\} ,$$

which after some minor manipulations becomes

$$(2.10) \quad \nu(\Gamma(\tfrac{\nu+1}{2}))^2 = 2(\Gamma(\tfrac{\nu+2}{2}))^2 \quad .$$

But this equality must fail for every $\nu \geq 0$. Indeed, if $\nu=0$, the left side of (2.10) vanishes, while the right side is 2 . If ν is positive, one side of (2.10) is an integer, while the other side is a rational multiple of π . Thus the assumption that F_ν has no zeros yields a contradiction, and Lemma 2.1 is proved. ∎

We can now give the

Proof of Theorem 2.2. Since for each pair (ν,n), the Padé numerator $Q_{\nu,n}(z)$ and Padé denominator $P_{\nu,n}(z)$ have no common factors, it suffices to show that for each fixed ν , the polynomials $P_{\nu,n}(z)$ have zeros of the form (2.4). Using the representation (1.3) the following integral formula can be derived:

$$(2.11) \quad (n+\nu)!\ P_{\nu,n}(z) = \int_0^\infty e^{-t}(t+z)^n t^\nu dt \quad , \quad (0 \leq t < +\infty) \quad .$$

Letting $z = n + \sqrt{n}\,w$ and making the change of variables $t=\sqrt{n}\,u$, $0 \leq u < +\infty$, in (2.11) we find that

$$(2.12) \quad (n+\nu)!\ P_{\nu,n}(n+\sqrt{n}\,w) = n^{\frac{2n+\nu+1}{2}} \int_0^\infty e^{-\sqrt{n}\,u}\{1 + \frac{w+u}{\sqrt{n}}\}^n u^\nu du.$$

The logarithm of the integrand above is, for u and w fixed and n large,

$$(\sqrt{n}\,w - \frac{w^2}{2}) - wu - \frac{u^2}{2} + \nu\ln u + O(\frac{1}{\sqrt{n}}) \quad ,$$

and so

$$\lim_{n\to\infty} e^{-\sqrt{n}\,u}\{1 + \frac{w+u}{\sqrt{n}}\}^n u^\nu / e^{\sqrt{n}\,w-w^2/2} = u^\nu e^{-wu-u^2/2} \quad .$$

Now the proof given by Newman and Rivlin [4] can be adapted here

to show, using the Lebesgue Dominated Convergence Theorem, that

$$(2.13)\qquad \lim_{n\to\infty} \frac{(n+\nu)!\ P_{\nu,n}(n+\sqrt{n}\ w)}{e^{\sqrt{n}\ w-w^2/2}\ n^{\frac{2n+\nu+1}{2}}} = \int_0^\infty u^\nu e^{-wu-u^2/2} du =: F_\nu(w) \quad ,$$

the convergence being uniform on compact subsets of the w-plane. Since, by Lemma 2.1, $F_\nu(w)(\not\equiv 0)$ has a finite zero, say w_ν, Hurwitz's Theorem implies that $P_{\nu,n}(n+\sqrt{n}\ w)$ possesses a zero, say $w=w_{\nu,n}$, such that $w_{\nu,n}\to w_\nu$ as $n\to\infty$. This means $P_{\nu,n}(z)$ has a zero of the form (2.4). ∎

Concerning pole-free <u>sectors</u> for the Padé approximants $R_{\nu,n}(z)$ the following is known:

THEOREM 2.3 (Saff, Varga [9], [11]). <u>For every</u> $\nu\geq 0$, $n\geq 2$, <u>the Padé approximant</u> $R_{\nu,n}(z)$ <u>for</u> e^{-z} <u>has no poles in the infinite sector</u>

$$(2.14)\qquad S_{\nu,n} := \{z : |\arg z| \leq \cos^{-1}\left(\frac{n-\nu-2}{n+\nu}\right)\} \quad .$$

<u>Furthermore, for any fixed</u> σ , $0<\sigma<+\infty$, <u>each element in the sequence of approximants</u> $\{R_{\nu_j,n_j}(z)\}_{j=1}^\infty$ <u>satisfying</u>

$$(2.15)\qquad \lim_{j\to\infty} n_j = +\infty\ ,\quad \lim_{j\to\infty} \nu_j/n_j = \sigma\ ,\quad \text{and}\quad \left(\frac{\nu_j+1}{n_j-1}\right) \geq \sigma\ ,$$

<u>for all</u> $j\geq 1$, <u>is pole-free in the infinite sector</u>

$$(2.16)\qquad S_\sigma := \{z : |\arg z| \leq \cos^{-1}\left(\frac{1-\sigma}{1+\sigma}\right)\} \quad ,$$

<u>and</u> S_σ <u>is the largest sector of the form</u> $|\arg z| \leq \mu$, $\mu>0$, <u>which is devoid of all poles of any sequence of approximants</u> $\{R_{\nu_j,n_j}(z)\}_{j=1}^\infty$ <u>satisfying</u> (2.15).

In particular, for any (fixed) $\sigma>0$, S_σ is the largest pole-free sector of the form $|\arg z| \leq \mu$, $\mu>0$, for the

sequence $\{R_{[\sigma n],n}(z)\}_{n=1}^{\infty}$, where $[\cdot]$ denotes the greatest integer function. This fact has an interesting geometric interpretation as explained in [9].

Using Theorems 2.2, 2.3, and the results in [11], we can deduce the following new result:

THEOREM 2.4. <u>A necessary and sufficient condition that a sequence of Padé approximants</u> $\{R_{\nu_k,n_k}(z)\}_{k=1}^{\infty}$, <u>with</u> $n_k \to \infty$, <u>be pole-free in some infinite sector</u> $|\arg z| < \mu$, $\mu>0$, <u>is that</u>

$$\liminf_{k\to\infty} \nu_k/n_k > 0 \ . \tag{2.17}$$

<u>Proof</u>. The sufficiency part follows immediately from Theorem 2.3. To prove necessity assume that (ν_k,n_k) is a sequence such that $n_k \to \infty$ and $\liminf_{k\to\infty} \nu_k/n_k=0$. Our aim is to show that for every $\mu>0$, there are infinitely many poles of the sequence $\{R_{\nu_k,n_k}(z)\}_{k=1}^{\infty}$ in the sector $|\arg z| < \mu$. For this purpose let $\{(\nu_j,n_j)\}_{j=1}^{\infty}$ denote a subsequence of $\{(\nu_k,n_k)\}_{k=1}^{\infty}$ for which $\lim_{j\to\infty} \nu_j/n_j=0$. We consider two separate cases:

<u>Case</u> 1: If some subsequence of $\{\nu_j\}_{j=1}^{\infty}$ is bounded, then there is evidently a subsequence $\{(\nu_\ell,n_\ell)\}_{\ell=1}^{\infty}$ of $\{(\nu_j,n_j)\}_{j=1}^{\infty}$ for which ν_ℓ is constant, say $\nu_\ell=\nu$ for all $\ell \geq 1$, and for which $\lim_{\ell\to\infty} n_\ell = \infty$. But as a consequence of Theorem 2.2, the sequence $\{R_{\nu,n_\ell}(z)\}_{\ell=1}^{\infty}$ has poles which asymptotically (as $n_\ell \to \infty$) lie on some parabola opening about the positive real axis. Therefore the sequence has infinitely many poles in any sector of the form $|\arg z| < \mu$, $\mu>0$.

<u>Case</u> 2: If $\nu_j \to \infty$ as $j\to\infty$, and $\lim_{j\to\infty} \nu_j/n_j=0$, then the result of Corollary 3.1 of [11] applied to the sequence $\{P_{\nu_j,n_j}(z)\}$ of Padé denominators for e^{-z} again shows that

there is no pole-free sector of the form $|\arg z| < \mu$, $\mu>0$, for the sequence $\{R_{\nu_j,n_j}(z)\}_{j=1}^{\infty}$. ∎

The next theorem has application to stability questions and extends results in [3] and [18].

THEOREM 2.5 (Saff, Varga [9]). *If* $n\leq\nu+4$, *the Padé approximant* $R_{\nu,n}(z)$ *for* e^{-z} *has all its poles in the open left half-plane*.

The above theorem is sharp in the sense that the approximant $R_{o,5}(z)$, for which $n=\nu+5$, does in fact have a pole in the right half-plane. However the following assertion can be made with regard to diagonal sequences of the table (1.4) of the form $\{R_{n-\tau,n}(z)\}_{n=\tau}^{\infty}$, $\tau\geq5$:

THEOREM 2.6 (Saff, Varga [9]). *For any integer* $\tau\geq5$, *there exists an integer* $m=m(\tau)$ *such that the approximants* $\{R_{n-\tau,n}(z)\}_{n=m}^{\infty}$ *have all their poles in the open left half-plane*.

3 Geometric convergence of Padé approximants in unbounded regions

In this section we discuss results concerning geometric convergence of Padé approximants on the nonnegative ray $[0,+\infty)$, and on infinite sectors of the form $|\arg z| \leq \mu$, $\mu>0$. First we set

$$(3.1)\quad \eta_{\nu,n}:=||e^{-x}-R_{\nu,n}(x)||_{L_\infty[0,+\infty)} \; .$$

Notice that when $\nu>n$, we have $\eta_{\nu,n}=|e^{-\infty}-R_{\nu,n}(\infty)| = \infty$. When $\nu\leq n$ the following estimates are known:

THEOREM 3.1 (Saff, Varga, Ni [12]). *For any nonnegative integers* ν *and* n *with* $0\leq\nu\leq n$, *there holds*

$$(3.2)\quad \frac{\gamma}{2^{n-\nu}\binom{n}{\nu}\,(n+1)^2} < \eta_{\nu,n} \le \frac{1}{2^{n-\nu}\binom{n}{\nu}} \quad ,$$

where γ *is a positive constant independent of* ν *and* n .

To state the next theorem we need the function $g(\beta)$ defined for $0\le\beta\le 1$ by

$$(3.3)\quad g(\beta):=\frac{\beta^{\beta}(1-\beta)^{1-\beta}}{2^{1-\beta}}\ ,\ 0<\beta<1\ ,\ g(0):=1/2\ ,\ g(1):=1\ .$$

THEOREM 3.2 (Saff, Varga, Ni [12]). *Let* $\{\nu(n)\}_{n=1}^{\infty}$ *be a sequence of nonnegative integers with* $0\le\nu(n)\le n$ *for all* n , *and satisfying* $\lim_{n\to\infty} \nu(n)/n = \beta$. *Then*

$$(3.4)\quad \lim_{n\to\infty} \eta_{\nu(n),n}^{1/n} = g(\beta) \quad .$$

As $\min_{0\le\beta\le 1} g(\beta) = g(1/3) = 1/3$, it follows from the above theorem that for *any* sequence $\{\nu(n)\}_{n=1}^{\infty}$, there holds

$$\liminf_{n\to\infty} \eta_{\nu(n),n}^{1/n} \ge \frac{1}{3} \quad ,$$

with equality possible for the sequence $\{R_{[n/3],n}(x)\}_{n=1}^{\infty}$. Indeed numerical computations appear to indicate that for each fixed n the smallest error $\eta_{\nu,n}$, $\nu=0,1,2,\ldots$, occurs when $\nu=[n/3]$. Another consequence of Theorem 3.2 is stated in

THEOREM 3.3 (Saff, Varga, Ni [12]). *A necessary and sufficient condition that a sequence of Padé approximants* $\{R_{\nu(n),n}(x)\}_{n=1}^{\infty}$ *converges geometrically in the uniform norm to* e^{-x} *on* $[0,+\infty)$ *is that*

$$(3.5)\quad \limsup_{n\to\infty} \frac{\nu(n)}{n} < 1 \quad .$$

Concerning geometric convergence in the uniform norm over infinite sectors we shall prove the following new result:

THEOREM 3.4. _A necessary and sufficient condition that a sequence of Padé approximants_ $\{R_{\nu(n),n}(z)\}_{n=1}^{\infty}$ _converges geometrically in the uniform norm to_ e^{-z} _in some infinite sector_ $S_\mu:=\{z:|\arg z| \leq \mu\}$, $\mu>0$, _is that_

$$(3.6) \quad 0 < \liminf_{n\to\infty} \frac{\nu(n)}{n} \leq \limsup_{n\to\infty} \frac{\nu(n)}{n} < 1 \quad .$$

Proof. That condition (3.6) is sufficient to ensure geometric convergence in some S_μ , $\mu>0$, is proved in [12]. To demonstrate necessity we assume that for some $\mu>0$, the sequence $\{R_{\nu(n),n}(z)\}_{n=1}^{\infty}$ satisfies

$$(3.7) \quad \limsup_{n\to\infty} ||e^{-z}-R_{\nu(n),n}(z)||_{L_\infty(S_\mu)}^{1/n} < 1 \quad .$$

Since S_μ contains the ray $[0,+\infty)$, it follows from Theorem 3.3 that $\limsup_{n\to\infty} \nu(n)/n < 1$. Furthermore, as (3.7) evidently implies that for n large enough, the poles of the sequence $\{R_{\nu(n),n}(z)\}$ must omit the sector S_μ , Theorem 2.4 implies $\liminf_{n\to\infty} \nu(n)/n > 0$. ∎

Concerning estimates for the size of the sector S_μ of geometric convergence for a sequence satisfying (3.6), the reader is referred to [12]. We remark that although no column of the table (1.4) converges geometrically to e^{-z} in an infinite sector, each column does, in fact, converge geometrically to e^{-z} on an unbounded parabolic region (see [10]).

Of course the poles of the Padé approximants to e^{-z} are, in general, not all real. For computational purposes it is sometimes desirable to deal with rational approximations whose poles are all real and coincident. In [7] it is shown that there exists a sequence of rational functions of the form

$$r_n(x) = \frac{p_{n-1}(x)}{(1+\frac{x}{n})^n}, \quad \deg p_{n-1} \leq n-1, \quad n=1,2,\ldots,$$

such that

$$||e^{-x} - r_n(x)||_{L_\infty[0,+\infty)} = \mathcal{O}(\frac{n}{2^n}) \quad \text{as} \quad n\to\infty .$$

Some further properties of this sequence are discussed in [7].

References

1 Boas, R.P., Entire Functions, Academic Press, Inc., New York, 1954.

2 Dieudonné, J., Sur les zeroes des polynomes-sections de e^x, Bull. Sci. Math., 70 (1935), 333-351.

3 Ehle, B.L., A-stable methods and Padé approximation to the exponential, SIAM J. Math. Anal., 4 (1973), 671-680.

4 Newman, D.J. and T.J. Rivlin, The zeros of partial sums of the exponential function, J. Approximation Theory, 5 (1972), 405-412.

5 Newman, D.J. and T.J. Rivlin, Correction: The zeros of the partial sums of the exponential function, J. Approximation Theory (to appear).

6 Perron, O., Die Lehre von den Kettenbrüchen, 3rd ed., vol. 2, B.G. Teubner, Stuttgart 1957.

7 Saff, E.B., A. Schönhage, and R.S. Varga, Geometric convergence to e^{-z} by rational functions with real poles, Numer. Math. (to appear).

8 Saff, E.B. and R.S. Varga, Zero-free parabolic regions for sequences of polynomials, SIAM J. Math. Anal. (to appear).

9 Saff, E.B. and R.S. Varga, On the zeros and poles of Padé approximants to e^z, Numer. Math. (to appear).

10 Saff, E.B. and R.S. Varga, Convergence of Padé approximants to e^{-z} on unbounded sets, J. Approximation Theory, 13 (1975), 470-488.

11 Saff, E.B. and R.S. Varga, On the sharpness of theorems concerning zero-free regions for certain sequences of polynomials, Numer. Math. (to appear).

12 Saff, E.B., R.S. Varga, and W.-C. Ni, Geometric convergence of rational approximations to e^{-z} in infinite sectors, Numer. Math. (to appear).

13 Szegö, G., Über eine Eigenschaft der Exponentialreihe, Berlin Math. Ges. Sitzungsber., 23 (1924), 50-64.

14 Underhill, C. and A. Wragg, Convergence properties of Padé approximants to exp(z) and their derivatives, J. Inst. Maths. Applics., 11 (1973), 361-367.

15 Van Loan, C., A study of the matrix exponential, Numerical Analysis Report No. 10, Univ. of Manchester, Manchester, England, 1975.

16 Van Rossum, H., On the poles of Padé approximants to e^z, Nieuw Archief voor Wiskunde (1) XIX (1971), 37-45.

17 Varga, R.S., Semi-infinite and infinite strips free of zeros, Rend. Sem. Mat. Univ. Politecnico Torino, 11 (1952), 289-296.

18 Wimp, J., On the zeros of a confluent hypergeometric function, Proc. Amer. Math. Soc., 16 (1965), 281-283.

E. B. Saff*
Department of Mathematics
University of South Florida
Tampa, Florida 33620

R. S. Varga**
Department of Mathematics
Kent State University
Kent, Ohio 44242

*Research supported in part by the Air Force Office of Scientific Research under Grant AFOSR-74-2688.

**Research supported in part by the Air Force Office of Scientific Research under Grant AFOSR-74-2729, and by the Energy Research and Development Administration (ERDA) under Grant E(11-1)-2075.

CALCULATION OF BEST APPROXIMATIONS BY RATIONAL SPLINES

R. Schaback

The purpose of this paper is to construct best Chebyshev approximations from a class γ of rational spline functions having fixed knots and free simple poles. As a first step, linear tangent spaces consisting of rational splines with fixed knots and free double poles are calculated. The next step is the proof of strong uniqueness properties of best approximations from γ and from the tangent spaces. This implies quadratic local convergence of a Newton-type algorithm by L. Cromme for calculating best approximations from γ. The numerical performance of the algorithm is discussed.

1 Introduction

Let $I = [a,b]$ be a compact interval of the real numbers R, divided into subintervals $I_j = [x_j, x_{j+1}]$ for $j = 1,\ldots,n$ by fixed "knots" $a = x_o < x_1 < \ldots < x_{n+1} = b$. We consider best approximations s to given functions $f \in C(I)$ with respect to the Chebyshev (uniform) norm on C(I). The approximations are taken from the set

$$(1) \quad \gamma = \{ s \in C^2(I) \mid s'' > o,\ s = \frac{p_j}{q_j} \text{ on } I_j,\ q_j \in P_1,\ p_j \in P_2 \}$$

of special rational spline functions [1,2,5,8], where P_k denotes the space of polynomials not exceeding k. The class of all rational spline functions is exactly the class of functions processed by computers. Thus (1) serves as a step towards the investigation of the best possible approximation of a function on a computer. Another argument for considering rational splines is the combination of the flexibility of splines and rational functions (see examples in [5,8].

Preceding page blank

The main result of this paper (Theorem 2 below) states that any locally strongly unique best approximation (for which a sufficient criterion is given) can be calculated by a locally quadratically convergent iteration involving only linear approximation problems. The concept of strong uniqueness is used here to ensure good numerical behavior; this is quite natural in view of the theorem of Freud [3], stating that strong uniqueness ensures Lipschitz continuous dependence of best approximations from the data.

2 Parametrization of γ

On the open set

$$A = \{ a = (y_o, M_o, \ldots, M_{n+1}, y_{n+1}) \in R^{n+4} \mid M_o, \ldots, M_{n+1} > o \}$$

in R^{n+4} we can parametrize γ by regarding vectors $a \in A$ as vectors of values $(s(x_o), \sqrt[3]{s''(x_o)}, \ldots, \sqrt[3]{s''(x_{n+1})}, s(x_{n+1}))$ of a spline $s = F(a) \in \gamma$. The explicit construction of s from a is carried out by evaluating (3.4) in [5], solving (2) for $y_1, \ldots, y_n$, and evaluating (3.7) in [5]. For fixed $a \in A$ we get the value $F_a'(b) = u$ of the Fréchet derivative of F at a with argument $b = (z_o, N_o, \ldots, N_{n+1}, z_{n+1}) \in R^{n+4}$ by the formulae

$$2\,D_j = (h_j N_{j-1} + h_j N_{j+1}) M_j^2 + 2 M_j N_j (h_j M_{j-1} + h_j M_{j+1}),$$

$$(2) \quad D_j = \frac{1}{h_{j-1} + h_j} \left(\frac{y_{j+1} - y_j}{h_j} - \frac{y_j - y_{j-1}}{h_{j-1}} \right) \quad (1 \leqq j \leqq n),$$

(which is to be solved for $y_1, \ldots, y_n$, setting $y_o = z_o$, $y_{n+1} = z_{n+1}$), and

$$u(t) = y_{j+1} + (t - x_{j+1}) \left[\frac{y_{j+1} - y_j}{h_j} + \frac{1}{2} h_j M_{j+1} (M_{j+1} N_j + 2 M_j N_{j+1}) \right] +$$

$$+ v_j(t) \left\{ 3 h_j M_j^2 N_{j+1} - (x_{j+1} - t)(3 M_j^2 N_{j+1} - 2 M_j M_{j+1} N_{j+1} - M_{j+1}^2 N_j) \right\},$$

$$v_j(t) = \frac{M_{j+1} (t - x_{j+1})^2}{2 h_j \left(1 - (x_{j+1} - t) \frac{M_j - M_{j+1}}{h_j M_j}\right)^2}, \quad h_j = x_{j+1} - x_j,$$

for $t \in I_j$, $j = o,\ldots,n$. Then

$$u(x_j) = y_j, \quad u''(x_j) = 3M_j^2 N_j, \quad j = o,\ldots,n+1.$$

These formulae are the derivatives of those defining F. Since for fixed a there are no difficulties concerning singularities, one can show that they actually describe the continous Fréchet derivative F'_a of F at a. These parametrizations define certain interpolating splines; nevertheless they are explicit except for solving a 3-banded linear system of equations.

3 Properties of the tangent spaces

We first recall the concept of strong uniqueness :

DEFINITION. An approximation v from a set V of elements of a normed space X is a (locally) strongly unique best approximation to a given element $f \in X$ iff there is a $K > o$ such that

$$||w - f|| \geqq ||v - f|| + K||w - v||$$

holds for all $w \in V$ (for all $w \in U \cap V$, where U is a neighborhood of u, respectively).

For $s = F(a)$ in the above sense, let $T_s = F'_a(R^{n+4})$ denote the tangent space to γ in s. We now list some properties of T_s:

THEOREM 1. Let s have denominators q_j in I_j, $j = o,\ldots,n$, and assume $q_j > o$ in I_j. Then, for $j = o,\ldots,n$:

(a) If $r_j \in P_3$, then $(r_j/q_j^2)''$ has at most one zero in I_j or vanishes identically.

(b) T_s coincides with the space of C^2-functions on I having the form r_j/q_j^2 in I_j, where $r_j \in P_3$.

(c) The Fréchet derivative F'_a defined above is an isomorphism between R^{n+4} and T_s.

(d) Let $g \in C(I)$ alternate in n+5 points $t_1 < t_2 < \ldots < t_{n+5}$ of I and assume

(3) $t_{i+1} < x_i < t_{i+4}$, $i = 1,\ldots,n.$

Then 0 is a strongly unique best approximation to g in T_s.

(e) Let the error function $g = f - s$ for $f \in C(I)$ and s have the properties assumed above.

Then s is a unique best approximation to f; in addition, s is locally strongly unique.

Indication of proof : Statement (a) follows by direct calculation. To show (b) and (c), one has to apply unicity arguments (see Satz 5 in [5]) for suitable interpolating functions provided by section 2. The proof of (d) proceeds along the lines of [6](see examples 6 and 7 there), while (e) follows from (d) by results of Braess and Werner ([2], Satz 2.4), and of Wulbert ([9], Lemma 9).

Remarks : (1) Condition (3) is satisfied for the approximation of generalized monosplines (Braess [1]).

(2) Condition (3) seems to be inevitable. This is indicated by the uniqueness properties of approximations by cubic splines [7], since these form T_s in case s is a parabola.

(3) The spaces T_s are generalizations of cubic splines; they can easily be generalized and seem worth investigating.

(4) The results of this section do not carry over to the closure $\bar{\gamma}$ of γ studied by Werner and Braess [2,8]. Even for n = 1 the characterization of locally strongly unique best approximations is rather complicated.

(5) Theorem 1 remains valid if I is replaced by a discrete set of points containing the knots and at least n+5 points.

4 Numerical procedure

We apply the following algorithm of Cromme [4] :

START : *Choose* $a^1 \in A$.

ITERATION : *Given* $a^i \in A$, *calculate* $F(a^i) = s_i \in \gamma$ *and a best approximation* $u_i \in T_{s_i}$ *to* $f - s_i$ *yielding a parameter vector* $b^i \in R^{n+4}$. *Form a new parameter vector* $a^{i+1} = a^i + \lambda b^i$ *with a suitable* $\lambda \in (o,1]$ *to ensure* $a^{i+1} \in A$.

We then have according to [4]

THEOREM 2. *Let the hypotheses of Theorem* 1,(e) *prevail. Then the above algorithm converges quadratically to* s *if it is started sufficiently near* s. *Furthermore, one can set* $\lambda = 1$ *in this case*.

Some remarks to this algorithm seem appropriate :

(1) The algorithm does not involve any nonlinear processes; it consists only of evaluations like those in section 2 and, of course, a sequence of linear approximation problems.

(2) The parameter λ is used to enlarge the convergence domain (by using a small λ whenever $\lambda=1$ would lead out of A) and to allow the creeping to the boundary of A.

(3) Of course one could seek an optimal λ in the way that $\|f - F(a^i + \lambda b^i)\|$ is minimized. On the other hand, one could split λ into n+4 components and adjust them separately to maintain $a^{i+1} \in A$. The performance of these variations of the algorithm is yet to be tested.

5 Numerical examples

Of course there are many examples showing the features of the algorithm stated above. As a typical case, the approximation of cosh(x) on [-1,+1] always converged quadratically and it never took more than 5 iterations to give a stable

13-digit precision. Several placements of up to 5 inner knots were tested, and the simple one-sided starting approximation $1 + 0.5x^2$ was used. The correct alternation behavior came up after one or two iterations.

So it appears to be better to comment on the shortcomings of the algorithm :

(1) There may be only local convergence. But for examples satisfying the conditions of Theorem 2 the algorithm seems to be globally convergent, though it often creeps along due to small values of λ when started far away from the best approximation.

(2) There may be no best approximation to f in γ. Then the algorithm moves very slowly to the boundary of γ; the convergence may even be poor when one is far away from the boundary. This is due to the fact that the approximations from T_s ignore the convexity needed for approximations from γ.

(3) There may be cases where f has a best approximation s in γ but (3) does not hold. We did not try specially constructed functions; for the commonplace functions we tested, either (3) came out to be satisfied or the situation (2) occurred.

References

1 Braess, D., Rationale Interpolation, Normalität und Monosplines, Numer. Math., 22 (1974), 219 - 232.

2 Braess, D., and H. Werner, Tschebyscheff - Approximation mit einer Klasse rationaler Splinefunktionen, II, J. Approximation Theory, 10 (1974), 379 - 399.

3 Cheney, E. W., Introduction to Approximation Theory, Mc-Graw-Hill, New York 1966.

4 Cromme, L., Eine Klasse von Verfahren zur Ermittlung bester nichtlinearer Tschebyscheff - Approximationen, Numer. Math. (to appear).

5 Schaback, R., Spezielle rationale Splinefunktionen, J. Approximation Theory, 7 (1973), 281 - 292.

6 Schaback, R., On Alternation Numbers in Nonlinear Chebyshev Approximation, J. Approximation Theory (to appear).

7 Schumaker, L. L., Uniform approximation by Tchebycheffian spline functions, J. Math. Mech., 18 (1968), 369 - 378.

8 Werner, H., Tschebyscheff - Approximation mit einer Klasse rationaler Spline - Funktionen, J. Approximation Theory, 10 (1974), 74 - 92.

9 Wulbert, D., Uniqueness and Differential Characterization of Approximations from Manifolds of Functions, Amer. J. Math., 93 (1971), 350 - 366.

R. Schaback*
Institut für Angewandte Mathematik
der Universität Bonn
D - 5300 Bonn
Federal Republic of Germany

*Supported in part by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 72.

The author wishes to thank Dipl. Math. H. Foerster for carrying out a series of numerical examples on the IBM 370/168 of the Regionales Hochschulrechenzentrum Bonn.

MAXIMUM NORM ERROR ESTIMATES IN THE FINITE ELEMENT METHOD FOR POISSON EQUATION ON PLANE DOMAINS WITH CORNERS

A. H. Schatz and L. B. Wahlbin

Estimates are given for the error in the L_∞-norm of finite element approximations of domains with corners.

In this lecture we shall discuss some particular cases of results which were recently obtained by the authors. Since the precise statements of these results are somewhat technical, we shall restrict ourselves here to a sketch. A more general and detailed account with proofs will be given elsewhere.

Let Ω be a bounded simply connected domain in the plane whose boundary $\partial\Omega$ consists of a finite number of straight line segments with interior angles $0 < \alpha_1 \leq \cdots \leq \alpha_M < 2\pi$, i.e. Ω is a polygonal domain which may have a number of slits. Consider Dirichlet's problem

(1) $$-\Delta u = f \quad \text{on } \Omega$$

(2) $$u = 0 \quad \text{on } \partial\Omega .$$

For simplicity, let f be smooth on $\bar{\Omega}$. In this case it is well known that u is smooth away from the corners but in general may be badly behaved near the corner points. Let us suppose that the vertex of the j^{th} corner is located at the origin (see Figure 1).

Then for R sufficiently small (where (R,θ) are polar coordinates) the "singular part" of u behaves like (c.f. [9])

Preceding page blank

(3) $$C_j R^{\frac{\pi}{\alpha_j}} \sin\left(\frac{\pi}{\alpha_j}\theta\right) (\ln R)^q,$$

where $q = 0$ if $\alpha_j \neq \pi/m$, $m > 2$ an integer, and $q = 1$ otherwise.

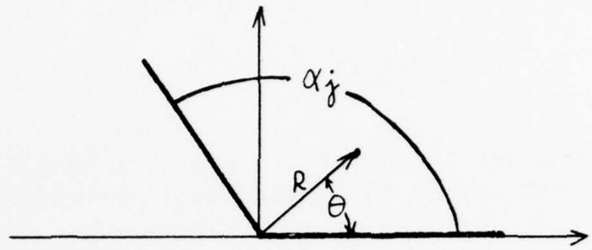


Figure 1

We wish to approximate u using the finite element method. To this end let $0 \leq h \leq 1$ be a parameter and let $S^h(\Omega)$ denote a family of finite dimensional subspaces of $W^1_\infty(\Omega) \cap \overset{\circ}{W}{}^1_2(\Omega)$. Here $W^m_p(\Omega)$ denotes the usual Sobolev space and $\overset{\circ}{W}{}^m_p(\Omega)$ is the closure of $C_0^\infty(\Omega)$ in the norm of $W^m_p(\Omega)$. The approximate solution $u_h \in S^h$ is then defined as the unique solution of the Ritz Galerkin Equations,

(4) $$D(u_h, \phi) \equiv \int_\Omega \sum_{i=1}^{2} \frac{\partial u_h}{\partial x_i} \frac{\partial \phi}{\partial x_i}\, dx = \int_\Omega f\phi\, dx, \quad \text{for all } \phi \in S^h,$$

or

(5) $$D(u - u_h, \phi) = 0 \quad \text{for all } \phi \in S^h.$$

We will investigate maximum norm estimates for the error $u(x) - u_h(x)$ as h tends to zero, where S^h is chosen to have properties which are shared by many subspaces used in practice. Rather than detail the abstracted properties we need, we confine ourselves here to listing two examples of classes of subspaces which satisfy our hypothesis. In what follows $r \geq 2$ is a given integer.

Example 1. For each $0 < h \leq 1$ let T_h be a quasi-uniform triangulation of Ω with h the maximum length of any side of any triangle $\tau \in T_h$. We define S^h as the subspace of $C^0(\bar{\Omega})$ consisting of those functions which vanish on $\partial\Omega$ and which, when restricted to each triangle τ, are polynomials of degree $r-1$.

Example 2. Suppose the boundary segments of the domain are parallel to the co-ordinate axis. Then it is often possible to define $S^h(\Omega)$ in the following way:

(i) $S^h(\Omega)$ is a subspace of the restriction to Ω of tensor products of piecewise polynomials.

(ii) The functions in $S^h(\Omega)$ vanish on $\partial\Omega$.

Roughly speaking, r is the optimal order of h which a function in $W^r(\Omega) \cap \overset{\circ}{W}{}_2^1(\Omega)$ can be approximated in the maximum norm by a function in S^h. With regard to very recent work on maximum norm estimates we refer the reader to Douglas, Dupont, and Wahlbin [3], Natterer [5], Nitsche [6], Scott [8] and Schatz and Wahlbin [7], where other relevant references may be found. Concerning the finite element method for domains with corners see Babuška and Rosenzweig [2] and Strang [9] for estimates in L_2 based norms.

Since (even for smooth f) the smoothness of the solution u may differ near each corner and again in the interior of Ω, we wish to investigate the error $u(x) - u_h(x)$ locally. To this end, let Ω_j, $j = 1,\cdots,M$, be any neighborhood (in $\bar{\Omega}$) of the j^{th} corner such that $\bar{\Omega}_j$ does not contain any other vertex of Ω and let Ω_0 be any subset of $\bar{\Omega}$ such that $\bar{\Omega}_0$ does not contain any vertex of Ω (i.e. avoids the corners). We then have the following error estimates:

THEOREM 1. Let $\varepsilon > 0$ be arbitrary but fixed. There exists a constant C such that for h sufficiently small the following hold:

(6) $$||u-u_h||_{L^\infty(\Omega_j)} \le Ch^{\min(\pi/\alpha_j, r, 2\pi/\alpha_M)-\varepsilon}, \quad j=1,\cdots,M,$$

and

(7) $$||u-u_h||_{L^\infty(\Omega_0)} \le Ch^{\min(r, 2\pi/\alpha_M)-\varepsilon}$$

To illustrate these estimates, let Ω be the L shaped region (see Figure 2) with $\alpha_1 = \cdots = \alpha_5 = \pi/2$ and

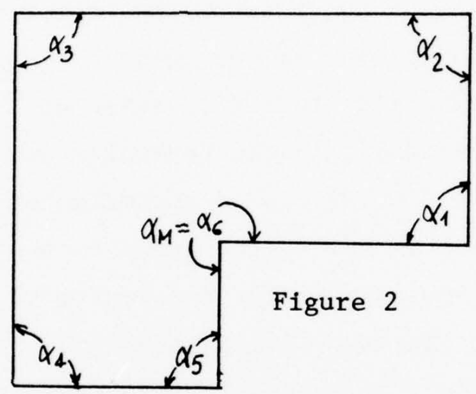


Figure 2

$\alpha_6 = \alpha_M = 3/2\pi$. Then for any of the subspaces S^h given in examples 1 and 2 we have that away from the re-entrant corner,

(8) $$||u-u_h||_{L^\infty(\Omega_j)} \le Ch^{4/3-\varepsilon} \quad j=0,1,\cdots,6.$$

In any neighborhood of the re-entrant corner

(9) $$||u-u_h||_{L^\infty(\Omega_6)} \le Ch^{2/3-\varepsilon}.$$

In this case the estimates (8) and (9) agree with those found in computer experiments shown to us by Babuška [1] (cf. also [9]). Let us note that the estimates (8) and (9) take into account the worst possible behavior of u near the corners. In the case of the L-shaped region Douglas, Dupont, and Wheeler

[4] have shown that if the solution u is sufficiently smooth everywhere on $\bar{\Omega}$ and if for S^h one takes the tensor products of continuous one-dimensional piecewise polynomials (a special case of example 2), then in maximum norm the rate of convergence is order h^r everywhere on $\bar{\Omega}$. We can show that if Ω is a convex polygon and S^h is taken, for example, to be the subspaces given in examples 1 and 2 then the convergence is of order $h^{4-\varepsilon}$ (for sufficiently smooth solutions).

Returning to the general case, let us now assume that the coefficient C_M in the "singular part" of u, (3) (associated with the maximum interior angle α_M) is different from zero. To simplify the discussion here, let us assume $\pi < \alpha_M < 2\pi$ and $\alpha_{M-1} < \alpha_M$. The estimates (6) and (7) then predict a lower rate of convergence near the M^{th} corner than elsewhere. It is easy to see that

$$||u - u_h||_{L_\infty(\Omega)} \geq ch^{\frac{\pi}{\alpha_M}+\varepsilon}$$

where $\varepsilon > 0$ is arbitrary and $c = c(\varepsilon) > 0$. This suggests that perhaps the maximum pointwise error must occur near the vertex of the maximum interior angle. Under the above assumptions we can prove the following:

<u>THEOREM 2</u>. <u>Let</u> $\delta > 0$ <u>be arbitrary but fixed. For each</u> h <u>let</u> $x_h \in \bar{\Omega}$ <u>be such that</u> $|u(x_h) - u_h(x_h)| = ||u - u_h||_{L^\infty(\Omega)}$. <u>Then for</u> h <u>sufficiently small</u>

$$|x_h - x_M| \leq h^{1-\delta} ,$$

<u>where</u> x_M <u>is the position of the vertex of the maximum interior angle</u>.

Finally, let us briefly mention some other results.

1) <u>Calculation of the coefficient</u> C_M <u>in (3)</u>. Let us assume

that $\pi < \alpha_M < 2\pi$. The calculation of a good approximation C_m^h to C_M of importance in some physical problems. For fixed $0 < \theta < \alpha_M$, one can define

$$C_M^h = \frac{u_h(d,\theta)}{d^{\pi/\alpha_M} \sin(\frac{\pi}{\alpha_M}\theta)} ,$$

where $d \approx h^{2\pi/(\alpha_M+\pi)}$. It can be shown that for h sufficiently small,

$$|C_M^h - C_M| \leq Ch^{\frac{2\Pi}{\alpha_M}(\alpha_M - \Pi/(\alpha_M + \Pi))}$$

2) Refinements. In order to improve the rate of convergence one may take a finer mesh near the corners. We have partial results in this direction.

3) Singular functions. Another way to increase the accuracy of the approximate solution is to add to S^h certain functions which mimic the behavior of the solution near the vertices. We also have partial results for this method.

References

1. Babuška, I, Personal communication.

2. Babuška, I. and M. B. Rosenzweig, A finite element scheme for domains with corners, Numer. Math., 20 (1972), 1-21.

3. Douglas, J. Jr., T. Dupont and L. B. Wahlbin, Optimal L_∞ error estimates for Galerkin approximations to solutions of two-point boundary value problems, Math. Comp., 29 (1975), 475-483.

4. Douglas, J. Jr., T. Dupont and M. F. Wheeler, An L^∞-estimate and a superconvergence result for a Galerkin method for elliptic equations based on tensor products of piecewise polynomials, to appear.

5. Natterr, F., Über die punktweise Konvergenz Finiter Elemente, to appear.

6. Nitsche, J. A., L_∞-convergence for finite element approximation, 2. Conference on Finite Elements, Rennes, France, May 12-14, 1975.

7. Schatz, A. H. and L. B. Wahlbin, Interior maximum norm estimates for finite element methods, to appear.

8. Scott, R., Optimal L^∞ estimates for the finite element method on irregular meshes, to appear.

9. Strang, G. and G. Fix, An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs 1973.

A. H. Schatz*
L. B. Wahlbin*
Department of Mathematics
Cornell University
Ithaca, N. Y. 14853

* Supported in part by the National Science Foundation

SOME INVERSE THEOREMS FOR BEST APPROXIMATION BY Λ-SPLINES

Karl Scherer

Some inverse theorems for best L_p-approximation by Λ-splines are presented, including those of a recent paper of Johnen/Scherer for sequences of (strongly) mixed and nested partitions of knots, as well as the corresponding one for the sequence of equidistant partitions and a weak saturation theorem for general sequences of partitions.

1 Introduction

The purpose of this note is to complete the discussion in [4] where the following inverse theorem was proved:

THEOREM 1. <u>Let</u> $\{\Delta_k\}_{k=1}^{\infty}$ <u>be a strongly mixed sequence of partitions of</u> $[a,b]$. <u>Then there exists a constant</u> $c>o$ <u>such that for all</u> $f \in L_p(a,b)$ <u>and all</u> $k \geq k_o$

$$(1)\quad K(\overline{\Delta}_k, f; p, \Lambda) \leq c \sup_{l \geq k} E_{\Delta_l}^{(\Lambda)}(f;p) \qquad (1 \leq p \leq \infty)$$

In case $p = \infty$ the assumption "strongly mixed" can be weakened to "mixed". Here the best approximation

$$(2)\quad E_{\Delta}^{(\Lambda)}(f;p) = \inf_{s \in Sp(\Lambda,\Delta)} \| f-s \|_p$$

is introduced, where $Sp(\Lambda,\Delta)$ denotes the class of Λ-spline-functions s which belong to the nullspace N_{Λ} of the linear differential operator

$$(3)\quad \Lambda = D^n + \sum_{j=o}^{n-1} a_j D^j \qquad (a_j \in C^j[a,b])$$

on each subsegment $I_i = (x_i, x_{i+1})$ $(i=o,\ldots,N-1)$ of a given partition

$$(4)\quad \Delta : a = x_o < x_1 < \ldots < x_{N-1} < x_N = b$$

with $\overline{\Delta} = \max_{1 \leq i \leq N} (x_i - x_{i-1})$, $\underline{\Delta} = \min_{1 \leq i \leq N} (x_i - x_{i-1})$. It should be

remarked that smoothness of the Λ-splines is not required since this would give no better results in (1).

The second quantity in (1) is defined for each $f \in L_p(a,b)$ and $o<t<\infty$ by

(5) $K(t,f;p,\Lambda) = \inf\{\|f-g\|_p + t^n\|\Lambda g\|_p : g \in W_p^n(a,b)\}$,

where $W_p^n(a,b)$ denotes the usual Sobolev space of order n. In view of the fact that (cf. [4])

(6) $\liminf_{t\to o} t^{-n}K(t,f;p,\Lambda) = 0 \Longrightarrow f \in N_\Lambda$,

inequality (1) establishes in particular that such a sequence $\{E_{\Delta_l}^{(\Lambda)}(f;p)\}_{l=1}^{\infty}$ is saturated with order $\overline{\Delta}_l^{\,n}$ (if $\overline{\Delta}_l \geqslant \overline{\Delta}_{l+1}$), that is that $E_{\Delta_l}^{(\Lambda)}(f;p) = O(\overline{\Delta}_l^{\,n})$ for $l\to\infty$ implies $f \in N_\Lambda$ or $E_{\Delta_l}^{(\Lambda)}(f;p) = 0$ for all l.

The condition "mixed" imposed on the sequence $\{\Delta_k\}_{k=1}^{\infty}$ requires that there is a number k_o such that for $k \geqslant k_o$ there holds

(7) $\sup_{l\geqslant k} \min_i \rho_i(k,l) \geqslant d\underline{\Delta}_k$

with a constant independent of i,k,l where $\rho_i(k,l) = \mathrm{dist}(t_i,\Delta_l)$ and t_i is the i-th knot of Δ_k.
Strongly mixed means that in addition the sup in (7) can be extended over only a finite number (fixed with respect to k) of l's.

It is easy to construct simple examples of strongly mixed sequences of partitions. However, it could not be shown in [4] that the sequence

(8) $\Delta_N^* = \{t_{i,N}^* = i/N, \quad o<i<N\}$

of equidistant partitions is strongly mixed, so Theorem 1 applies to this case only for $p=\infty$. In this paper it will be shown that (the unit interval (0,1) is considered)

(9) $K(1/N,f;p,\Lambda) \leqslant \sup_{N\leqslant M\leqslant 2N} E^{(\Lambda)}_{\Delta^*_M}(f;p).$

A still sharper result, involving on the right side only $M = N, N+1$, can be proved for polynomial splines or, more generally, for Λ with constant coefficients a_j (see [5]).

2 Proof of the results

One begins with the inequality

(10) $K(t,f;p,\Lambda) \leqslant \|f-s_N\|_p + K(t,s_N;p,\Lambda)$

where s_N is an element of $S_p(\Delta^*_N,\Lambda)$ of best approximation to $f \in L_p(0,1)$ (for convenience the case $(a,b) = (0,1)$ is considered). By a lemma in [4] one can further estimate

$$K(t,s_N;p,\Lambda) \leqslant C\{\sum_{i=1}^{N} (1+t^n\varepsilon^{-np})\varepsilon \max_{o\leqslant k\leqslant n-1} |\varepsilon^k[s_N]^k_i|^p\}^{1/p}$$

with a constant c depending only on Λ and p. The $[s]^k_i$ are the jumps of the k-th derivative of s at the mesh points x_i and ε is a positive number $< \underline{\Delta} \leqslant 1$. Thus, taking $\varepsilon = t = 1/N$, one has

$$K(1/N,s_N;p,\Lambda) \leqslant C\{\sum_{i=1}^{N} \max_{o\leqslant k\leqslant n+1} N^{-(kp+1)}|[s_N]^k_i|^p\}^{1/p}.$$

By a Markov-type inequality for elements of $N(\Lambda)$ on arbitrary subintervals of $(0,1)$ one obtains for $N<M\leqslant 2N$

$$|[s_N]^k_i = |[s_N - s_M]^k_i| \leqslant c\,\rho_i(N,M)^{-k-1/p}\|s_N-s_M\|_{p,i},$$

where $\rho_i(N,M) = \mathrm{dist}(t^*_{i,N},\Delta^*_M)$ and $\|\cdot\|_{p,i}$ denotes the norm with respect to the subinterval $(t^*_{i,N} - \rho_i(N,M),\ t^*_{i,N}+\rho_i(N,M))$ Thus it follows

(11) $K(1/N,s_N;p,\Lambda) \leqslant$

$$C \max_{o\leqslant k\leqslant n-1} \{\sum_{i=1}^{N} N^{-(kp+1)}\ \rho_i(N,M)^{-(kp+1)}\|s_N - s_M\|_{p,i}\}^{1/p}$$

Now one splits up the sum into three parts,

(12) $$\sum_{i=1}^{N} = \sum_{i=1}^{\leqslant N/3} + \sum_{i\geqslant N/3}^{\leqslant 2N/3} + \sum_{i\geqslant 2N/3}^{N} \equiv \Sigma_1 + \Sigma_2 + \Sigma_3,$$

say. To handle the first sum one uses a mixing property, observed by Butler/Richards [2],

$$(13)\quad N^{-(\beta p+1)} \leq C \min_{o<t^*_{j,N}<\delta} \frac{1}{N} \sum_{M=N+1}^{2N} \varrho_j(N,M)^{\beta p+1}$$

for all $N \geq N_o$ and $\beta > o, o<\delta<1/2$, the constant C depending only on δ and β. Inserting this with $\beta = k+1$ gives

$$(14)\quad \Sigma_1 \leq C \sum_{i=1}^{\leq N/3} (1/N) \sum_{M=N+1}^{2N} N^p \varrho_i(N,M)^p \| s_N - s_M \|^p_{p,i}$$

$$\leq (C/N) \sum_{M=N+1}^{2N} \| s_N - s_M \|^p_p \leq C \max_{N\leq M\leq 2N} \| s_N - s_M \|^p_p .$$

Inequality (13) is also valid for all $t^*_{j,N} \in (1-\delta,1)$ by symmetry so that the same estimate as (14) also holds for Σ_3.

Concerning the middle term Σ_2 one proceeds somewhat different. In (11) one chooses $M = N + 1$ and verifies readily that $\varrho_i(N,N+1) \geq 1/3(N+1)$, so that one obtains, replacing the sum in (15) by the sum over $N/3 \leq i \leq 2N/3$,

$$(15)\quad \Sigma_2 \leq C \{ \sum_{i\leq N/3}^{\leq 2N/3} \| S_N - S_M \|^p_{p,i} \}^{1/p} \leq C \| S_N - S_{N+1} \|_p .$$

Combining (10)-(12) and (14), (15) gives the desired inequality (9).

Together with a direct theorem on best approximation by Λ splines (cf. [4]) it is now possible to state the following approximation theorem

THEOREM 2. *Let* $\varrho(t)$ *be a non-decreasing function on* $(0,1)$ *with* $\lim_{t\to o} \varrho(t) = 0$. *Then for* $f \in L_p(0,1), 1\leq p\leq\infty$, *there holds*

$E_N^{(\Lambda)}(f;p) = O[\varrho(1/N)], N\to\infty$, *if and only if* $K(t,f;p,\Lambda) = O[\varrho(t)],\ t\to 0$.

In particular, a saturation theorem holds for $\varrho(t) = O(t^n), t\to 0$, in view of (6). For the other nontrivial choices of $\varrho(t)$, a characterization of the above assertions in terms of Lipschitz spaces can be obtained (cf. [4]).

For general sequences of partitions a unified treatment of inverse theorems is not possible. In particular the case of sequences of nested partitions shows quite different features. However, if one assumes sufficient smoothness of the function to be approximated, a saturation theorem can be proved (the special case of nested partitions has been considered in [4].

THEOREM 3. <u>Let</u> $f \in W_p^n(a,b), 1<b\leq\infty$, <u>and</u> $\{\Delta_k\}_{k=1}^{\infty}$ <u>a sequence partitions of</u> $[a,b]$. <u>Then</u> $E_{\Delta_N}^{(\Lambda)}(f;p) = O(\Delta_N^n), N\to\infty$, <u>implies</u> $f \in N(\Lambda)$.

<u>Proof</u>. According to (10) and lemma in [4] one has

$$(16)\quad K(t,f;p,\Lambda)\leq E_{\Delta_N}^{(\Lambda)}(f;p)+C\{\sum_{i=1}^{N}(1+t^n\varepsilon^{-np})\varepsilon \max_{0\leq k\leq n-1}|\varepsilon^k[s]_i^k|^p\}^{1/p}$$

for $\varepsilon<\Delta$, where s is an element of best approximation to f. In order to estimate the jumps $[s]_i^k$ at the point x_i we need the fact that for each subinterval (e,d) of (a,b) there holds the Taylor formula

$$(17)\quad f(x) = u_e(x) + \int_e^b \hat{v}(x,\xi)\Lambda f(\xi)d\xi$$

with $u_e \in N(\Lambda)$ satisfying $u_e^{(j)}(e) = f^{(j)}(e), 0\leq j\leq n-1$ and the kernel satisfying $\hat{v}(x,\xi) = \sum_{k=1}^{n} u_k(\xi)u_k(x)$ for $a\leq\xi\leq x\leq b$ and $\hat{v}(x,\xi) = 0$ for $a\leq x\leq\xi\leq b$. By the Taylor expansion of $\hat{v}(x,\xi)$ (cf. [3]) it follows for $x \in (e,d)$

$$D_x^j\hat{v}(x,e) = (x-e)^{n-j-1}/(n-j-1)! + \mathcal{O}((x-e)^{n-j}).$$

Hence one obtains readily, using Hölder's inequality, that

$$(18)\quad \{\int_e^d |f(x)-u_e(x)|^p dx\}^{1/p} \leqslant C|d-e|^n\{\int_e^d |\Lambda f(\xi)|^p d\xi\}^{1/p}$$

with a constant not depending on f or e,d. Now, denoting by $u_i(x)$ the element of $N(\Lambda)$ in the sense of (17) at $e = x_i-\varepsilon$, one obtains by a Bernstein-Type inequality and (17)

$$\begin{aligned}\varepsilon^{kp+1}|[s]_i^k|^p &\leqslant 2^p\varepsilon^{kp+1}\{\sup_{x\in(x_i-\varepsilon,x_i+\varepsilon)}|s^{(k)}(x)-u_i^{(k)}(x)|\}^p\\ &\leqslant C\|s-u_i\|_{p,i}^p\\ &\leqslant C\{\|s-f\|_{p,i}+\varepsilon^n\|\Lambda f\|_{p,i}\}^p\end{aligned}$$

where $\|\cdot\|_{p,i}$ denotes the L_p-norm with respect to $(x_i-\varepsilon,x_i+\varepsilon)$.

Inserting this into (16) it follows for any $1\leqslant q<p\leqslant\infty$

$$\begin{aligned}K(t,f;q,\Lambda) &\leqslant C\{1+t^n\varepsilon^{-n}\}E_{\Delta_N}^{(\Lambda)}(f;q)+Ct^n\{\sum_{i=1}^N\|\Lambda f\|_{q,i}^q\}^{1/q}\\ &\leqslant C\{1+t^n\varepsilon^{-n}\}E_{\Delta_N}^{(\Lambda)}(f;p)+Ct^n\varepsilon^{1/q-1/p}\|\Lambda f\|_p .\end{aligned}$$

Now let $E_{\Delta_N}^{(\Lambda)}(f;p) = \underline{\Delta}_N^n\cdot\psi(N)$ with some positive function, $\psi(N)\to 0$ for $N\to\infty$. Then, for N sufficiently large, take $\varepsilon = \psi(N)^{1/2n}\underline{\Delta}_N$ so that

$$\underline{\Delta}_N^{-n}K(\underline{\Delta}_N,f;p,\Lambda) \leqslant C\{\psi(N)+\psi(N)^{1/2}\}+(\psi(N)^{1/2n}\underline{\Delta}_N)^{1/q-1/p}.$$

But by a lemma in [4] this implies $f\in N(\Lambda)$.

Theorem 3 generalizes a corresponding result in [1,p,174] for polynomial splines and $p=\infty$. It shows that functions which are to be approximated by Λ-splines with arbitrary high order of convergence cannot be found within the class

$W_p^n(a,b)$ of smooth functions. On the other hand, Theorem 3 in [4] implies - at least for sequences of nested partitions - that in case of smooth splines in $W_p^m(a,b)$, $1 \leq m \leq n-1$ such functions of high order of approximation must lie in the class $W_p^m(a,b)$.

In summary, Theorems 1-3 show that, concerning the order of best approximation, the situation for polynomial splines and the more general Λ-splines is the same. This fact is already heuristically explained by the fact that $\Lambda = a_n D^n + (\ldots), a_n > o$, is a "perturbation" of D^n.

References

1 Ahlberg, J.H., Nilson, E.N. and J.L. Walsh, The theory of splines and their applications, Academic Press, NY (1967).

2 Butler, G. and F. Richards, An L_p-saturation theorem for splines, Canad.J.Math. 24 (1972), 957-966.

3 Jerome, J.W., On uniform approximation by certain generalized spline functions, J. Approximation Theory 7 (1973), 143-154.

4 Johnen, H. and K. Scherer, Direct and inverse theorems for best approximation by Λ-splines, in Spline Functions, Karlsruhe, 1975, Lecture Notes in Mathematics 501, Springer Verlag, 1976, 116-131.

5 Scherer, K., Characterization of general Lipschitz classes by approximation with splines, SIAM J.Numer.Anal. 11 (1974), 283-304.

K. Scherer
Institut für Angewandte Mathematik der Universität Bonn
Wegelerstraße 6, 5300 Bonn

A GENERALIZATION OF A RESULT OF SUBBOTIN

A. Sharma and J. Tzimbalario

The object of this note is to generalize a result of Subbotin and find the relation between some linear differential operators and the corresponding generalized differences.

1 Introduction

Let n be a fixed positive integer and let $\{y_k\}_{-\infty}^{\infty}$ be a given sequence of real numbers having bounded n-th order differences. About ten years ago, Subbotin [4] proposed and solved the problem of finding a function $f(x)$ on the whole real line which interpolates the data $\{y_k\}$ at the integers and has a least n-th derivative in the sup-norm.

Our aim is to study a corresponding problem when the n-th derivative is replaced by certain linear differential operator $L_n(D)$ and the n-th order differences by suitable divided differences. Following a remark of Schoenberg [2] "for an interesting and viable theory, we require also translation invariance" and so we restrict ourselves to differential operators with constant coefficients such that its characteristic equation has only real roots. In this case the extremal function turns out to be a cardinal L-spline which has been studied recently by Micchelli [1] and Schoenberg [2],[3].

In §2, we introduce some notations and formulate the problem. We also give some properties of the zeros of certain polynomials which are needed in the proof of our main result. We omit the details of the proof.

Preceding page blank

2 Preliminaries

Let $0 = \gamma_0, \gamma_1, \cdots, \gamma_n$ be real constants and let $L_n(D)$ and $L^*_{n+1}(D)$ be given by

(2.1) $$L_n(D) = \prod_1^n (D-\gamma_j), \quad L^*_{n+1}(D) = D\, L_n(D), \quad D \equiv \frac{d}{dx}.$$

For a given sequence $\{y_k\}_{-\infty}^{\infty}$, set

(2.2) $$\Delta_L^n\, y_m = \prod_1^n (E-e^{\gamma_j})y_m, \quad \Delta_{L^*}^{n+1} y_m = (E-1)\Delta_L^n\, y_m$$

where $Ey_m = y_{m+1}$. If a real function $f(x)$ defined on the real line satisfies the condition $\sup_x |L_n(D)f| \le M$, then it can be easily seen that $\sup_m |\Delta_L^n\, f(m)|$ is bounded. We shall be concerned in the sequel with the following problem:

PROBLEM. *For any given sequence* $\{y_k\}_{-\infty}^{\infty}$ *satisfying the condition*

(2.3) $$\sup_{m \in Z} |\Delta_L^n\, y_m| \le 1,$$

find a function $f(x)$ *satisfying the following conditions*:

(2.4) $$f(k) = y_k, \quad k \in Z$$

(2.5) $$||L_n(D)f||_\infty = \sup_x |L_n(D)f(x)| \text{ is minimum}$$

when f *runs over the functions satisfying* (2.4).

For the solution of this problem we shall need some auxiliary functions and their properties. Let $\phi_n(x)$ be the unique solution of $L_n(D)y = 0$ satisfying the conditions

(2.6) $$\phi_n^{(\nu)}(0) = \delta_{\nu,n-1}, \quad \nu = 0,1,\cdots,n-1.$$

Set

(2.7) $$\psi_n(x) = \int_0^x \phi_n(t)dt.$$

Then $\psi_n(x)$ is the unique solution of $L^*_{n+1}(D)y = 0$ determined by the conditions $\psi_n^{(\nu)}(0) = \delta_{\nu n}$ $(\nu = 0,1,\cdots,n)$.

Since $\Delta_L^n e^{\gamma_j x} = 0$, $j = 1,\cdots,n$, it follows easily that

$$(2.8)\quad \Delta_L^n f(x) = \int_0^\infty \Delta_L^n \left(\phi_n(x-t)_+\right) L_n(D)f(t)dt$$

where Δ_L^n applies to the variable x and

$$(2.9)\quad \phi_n(x-t)_+ = \phi_n(x-t), \text{ for } t \le x, \text{ and } = 0 \text{ otherwise.}$$

We now introduce the function $\Pi_n^*(x;\alpha)$ where

$$(2.10)\quad \Pi_n^*(x;\alpha) = \sum_{\ell=-\infty}^{\infty} x^\ell \, \Delta_{L^*}^{n+1} \, \psi(\ell-n-\alpha)_+ \,.$$

As a function of α, $\Pi_n^*(x;\alpha)$ is a cardinal L^*-spline and as a function of x, it is a polynomial of degree n when $0 < \alpha < 1$ and is of degree $n-1$ when $\alpha = 0$. It is clear that

$$(2.11)\quad \Pi_n^*(x;\alpha+1) = x\,\Pi_n^*(x;\alpha).$$

In the language of Schoenberg, $\Pi_n^*(x;\alpha)$ is the "exponential L^*-spline of basis x". Following a result of Micchelli ([1] Theorem 2.3 and Coro. 2.4), it turns out that for a given α, $0 \le \alpha < 1$, all the zeros of $\Pi_n^*(x;\alpha)$ are simple and negative. (For a different proof of this we refer to Schoenberg [2]). Micchelli showed that

$$(2.12)\quad A_n^*\left(1-\alpha;\, \frac{1}{x}\right) = (-1)^{n-1} x\, A_n^*(\alpha;x).$$

Here $\Pi_n^*(x;\alpha) = A_n^*(\alpha;x) \prod_{j=o}^{n} (e^{\lambda_j}-x)$. From this it is easy to verify that

$$(2.13)\quad \begin{aligned} \Pi_n^*\left(\frac{1}{x};\, \frac{1}{2}\right) &= x^n\, \Pi_n^*\left(x;\, \frac{1}{2}\right), \quad n \text{ even} \\ \Pi_n^*\left(\frac{1}{x};\, 0\right) &= x^{n-1}\, \Pi_n^*(x;\, 0), \quad n \text{ odd.} \end{aligned}$$

It follows from this that for n even $\Pi_n^*(-1;0) = 0$ and for n odd $\Pi_n^*\left(-1; \frac{1}{2}\right) = 0$. Since $\Pi_n^*(-1;\alpha)$ has only one simple zero in $[0,1)$ ([1],[2]), it follows that

$$(2.14) \qquad \begin{aligned} &\Pi_n^*\left(-1; \tfrac{1}{2}\right) \neq 0, \quad n \text{ even} \\ &\Pi_n^*(-1;0) \neq 0, \quad n \text{ odd.} \end{aligned}$$

We shall also require later the exponential polynomial $P_n(v)$ given by

$$(2.15) \qquad P_n(v) = \sum_{j=-\infty}^{\infty} (-1)^j \, \Delta_L^n \, \phi_n(-j+1-v)_+ \, .$$

Comparing (2.15) with (2.10), we see that the polynomial $P_n(v)$ corresponds to the operator L and that $P_n(v) = \Pi_n(-1;v)$ in analogy with the notation of (2.10). It follows mutatis mutandis that

$$(2.16) \qquad P_n(1-v) = (-1)^{n+1} P_n(v).$$

Also the only zero of $P_n(v)$ in $[0,1)$ is a simple zero at $\frac{1}{2}$ when n is even and is a simple zero at 0 when n is odd.

When $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$, the exponential polynomial $P_n(v)$ reduces to the Euler polynomial except for a constant factor. It follows from a continuity argument and from a property of the Euler polynomials that

$$(2.17) \qquad |P_n(v)| = (-1)^{\left[\frac{n+1}{2}\right]} P_n(v), \quad \frac{1}{2} \leq v \leq 1.$$

3 Main Result

We shall give a complete answer to the Problem when the operator $L_n(D)$ is formally self-adjoint, that is, $L_n(-D) = (-1)^n L_n(D)$. In other words the sets $\{\gamma_1,\cdots,\gamma_n\}$ and $\{-\gamma_1,\cdots,-\gamma_n\}$ are equal. We then have the following

THEOREM 1. Let the numbers $\gamma_1,\cdots,\gamma_n$ be such that the operator $L_n(D)$ is formally self-adjoint. Then for each sequence $\{y_k\}_{-\infty}^{\infty}$ satisfying (2.4), there is a function $f(x)$ satisfying (2.5) such that

$$\sup_x |L_n(D)f(x)| \le A_{n,L}.$$

Here

$$A_{n,L} = \frac{1}{\int_0^1 |P_n(v)|dv} = \frac{(-1)^{[\frac{n}{2}]}}{\Pi_n^*(-1;\alpha)}$$

where $\alpha = \frac{1}{2}$ when n is even, and $\alpha = 0$ when n is odd. Here $P_n(v)$ and $\Pi_n^*(-1,\alpha)$ are given by (2.15) and (2.10) respectively. The constant $A_{n,L}$ is sharp.

In the special case when $\gamma_1 = \gamma_2 = \cdots = \gamma_n = 0$, the result was proved by Subbotin [4]. In order to prove this we first show that

$$A_{n,L} \ge \frac{1}{\int_0^1 |P_n(v)|dv}$$

and then prove that $A_{n,L} \le \dfrac{(-1)^{[\frac{n}{2}]}}{\Pi_n^*(-1;\alpha)}$. Finally we check that the two bounds are equal. The details will appear elsewhere.

4 Conclusion

In the above the lower estimate for $A_{n,L}$ remains valid even when the operator $L_n(D)$ is not formally self-adjoint.

The first non-trivial case occurs when $n = 2$. In this case using (2.10), it follows easily that if $L(D) = D^2 - \gamma^2$, then $8 \cosh \frac{\gamma}{2} \sinh^2 \frac{\gamma}{4} A_{2,L} = \gamma^2$. The exponential polynomial $P_n(v)$ can be looked upon as a generalization of the Euler polynomials $E_{n-1}(v)$ to which it reduces when $\gamma_1 = \gamma_2 = \cdots =$

$\gamma_n = 0$ except for a constant factor. It would be interesting to know if other properties of Euler polynomials can be carried over to $P_n(v)$.

It is possible to propose a more general problem by replacing (2.4) and (2.6) by inequalities in ℓ_p and L_p norm respectively. It appears that the solution to the more general problem can be obtained on the lines of Subbotin ([5],[6]), in the light of the above result.

References

1 Schoenberg, I.J., On Charles Micchelli's Theory of Cardinal L-Splines. MRC Report #1511 (1975). Studies in Splines and Approx. Theory, Academic Press, 1976.

2 Micchelli, C.A., Cardinal L-Splines. Studies in Splines and Approx. Theory. Academic Press, 1976.

3 Schoenberg, I.J., Cardinal Spline Interpolation. CBMS Regional Conference. Monograph No. 12, S.I.A.M. Philadelphia, 1973.

4 Subbotin, Ju. N., On the relation between finite differences and the corresponding derivatives. Proc. Steklov Inst. Math. 78(1965), 23-42.

5 Subbotin, Ju. N., Interpolation by functions with n-th derivative of minimum norm. Proc. Steklov Inst. Math. 88(1967), 31-65.

6 Subbotin, Ju. N., Approximation by Spline functions and estimates of diameters. Proc. Steklov. Inst. Math. 109(1971), 39-67.

A. Sharma
Department of Mathematics
University of Alberta
Edmonton, Alberta T6G 2G1
Canada

J. Tzimbalario
Department of Mathematics
University of Alberta
Edmonton, Alberta T6G 2G1
Canada

ON THE SMOOTHNESS OF LOCAL BEST L_p SPLINE APPROXIMATIONS

Philip W. Smith

This paper is concerned with best and local best L_p approximation by splines with variable knots. In particular it will be shown that local best L_p, $1 \leq p < \infty$, spline approximation to a continuous function always yields a continuous function. Such is not the case in L_∞. Conditions are also obtained on the nature of the possible discontinuities when approximating certain discontinuous functions. These results extend and amplify those in [3, 4] and [2].

1 Introduction

The notation we employ here will be easily recognized by those familiar with spline theory. Let a positive integer k be given and let $\underline{t}$ denote a knot vector satisfying $\underline{t} : 0 = t_1 = \ldots = t_k < t_{k+1} \leq \ldots \leq t_n < t_{n+1} = \ldots = t_{n+k} = 1$ with $t_{j+k} > t_j$ for $j = 1,\ldots, n$. The normalized B-spline $N_{i,k}(\underline{t},\cdot)$ corresponding to the knot vector $\underline{t}$ is defined by $N_{i,k}(t,\tau) = (t_{i+k}-t_i)\,[t_i,\ldots,t_{i+k}]\,(\cdot-\tau)_+^{k-1}$, $1 \leq i \leq n$ (cf. [1]). We denote the set of splines of order k with $n-k$ or fewer knots (counting multiplicities) in (0,1) by S_n^k. It is well known [1] that $s \in S_n^k$ if and only if $s(\cdot) = \sum_{i=1}^{n} A_i N_{i,k}(\underline{t},\cdot)$ for some knot vector $\underline{t}$ and some coefficients A_i.

Given a function $f \in L_p[0,1]$, we will call $s \in S_n^k$ a best $L_p[0,1]$ approximation to f if $||s-f||_p = \inf\{||y-f||_p : y \in S_n^k\}$. Similarly we will call $s \in S_n^k$ a local best $L_p[0,1]$ approximation to f if for some

$\varepsilon > 0$, $||f-s||_p = \inf\{||f-y||_p : y \in S_n^k$ and $||y-s||_p < \varepsilon\}$.

2 Main results

Let $f \in L_p[0,1]$ and $s(\underline{t},\cdot) = \sum_{j=1}^{n} A_j N_{j,k}(\underline{t},\cdot)$.
We will denote the error between f and s by $e(\underline{t},\tau) = f(\tau) - s(\underline{t},\tau)$. For any function g with left and right limits at α we set $[g](\alpha) \equiv g(\alpha+) - g(\alpha-)$. The first theorem relates discontinuities in a local best approximating spline to discontinuities in the error.

THEOREM 2.1. Let $s(\underline{t},\cdot) \in S_n^k$ be a local best $L_p[0,1]$ approximation to $f \in L_p[0,1]$, $1 \leq p < \infty$. Assume further that for some $\alpha \in (0,1)$, $\lim_{t\to\alpha+} f(t)$ and $\lim_{t\to\alpha-} f(t)$ exist. Then

$$\text{(2.1)}\quad \text{i)}\ e(\underline{t},\alpha-)\ [s(\underline{t},\cdot)]\ (\alpha) \leq 0 \ .$$
$$\text{ii)}\ e(\underline{t},\alpha+)\ [s(\underline{t},\cdot)]\ (\alpha) \geq 0 \ .$$

Note that (2.1) is trivial unless α is a point of discontinuity of s. Thus we may assume that $t_{m-1} < \alpha = t_m = \ldots = t_{m+k-1} < t_{m+k}$ where the $\{t_i\}$ are the components of the knot vector $\underline{t}$ for s. The idea of the proof is to differentiate the error functional, $||e(\underline{t},\cdot)||_p^p$, with respect to t_m and t_{m+k-1}. Since it is only possible to compute one-sided derivatives of this functional with respect to t_m and t_{m+k-1} we obtain the inequalities in (2.1). These results for $1 < p < \infty$ under only slightly more restrictive hypotheses were obatined in [2]. The case $p = 1$ is new.

It is now possible with the aid of Theorem 2.1 to obtain results concerning the continuity of local best spline approximations.

THEOREM 2.2. *If* $1 \le p < \infty$ *and* f *is continuous then any local best* $L_p\,[0,1]$ *approximation from* S_n^k *to* f *is continuous.*

Proof. Assume to the contrary that f has a discontinuous local best L_p approximation $s(\underline{t},\cdot)$ with a discontinuity at t_m. Without loss of generality we may assume that $[s(\underline{t},\cdot)]\,(t_m) < 0$. This implies via (2.1) that $e(\underline{t},t_m-) \ge 0$ and $e(\underline{t},t_m+) \le 0$. But then $0 \le e(\underline{t},t_m-) - e(\underline{t},t_m+) = [s(\underline{t},\cdot)]\,(t_m)$ which is a contradiction and completes the proof.

Setting $[a,b]$ to be the comvex hull of a and b we may now state a result on discontinuous spline approximation.

THEOREM 2.3. *Let* $s(\underline{t},\cdot)$ *be a local best approximant to* f *from* S_n^k. *If* s *is discontinuous at* t_m *then* $[s(\underline{t},t_m-)\ ,\ s(\underline{t},t_m+)] \subset [f(t-),\ f(t+)]$.

The proof follows quickly from Theorem 2.1.

References

1 de Boor, Carl, On Calculating with B-splines, J. Approximation Theory, 6 (1972), 50-62.

2 Chui, Charles K., Philip W. Smith, and Joseph D. Ward, On the Smoothness of Approximations from Nonlinear Spline Manifolds, (To appear).

3 Schumaker, L.L., Uniform Approximation by Chebyshev Spline Functions. II: Free Knots. SIAM J. Numer. Anal., 5 (1968), 647-656.

4 Schumaker, L.L., On the Smoothness of Best Spline Approximation, J. Approximation Theory, 2 (1969), 410-418.

Philip W. Smith*
Department of Mathematics
Texas A&M University
College Station, Texas 77843

*Work supported by the U. S. Army Research Office under Grant Number DAHC04-75-G-0186.

APPROXIMATION BY REGULAR SPLINES WITH FREE KNOTS

H. Werner

In this paper the concept of regular splines $S_{m,k}$ of degree k with m-1 interior knots is introduced and a survey of recent results on this subject is given. Then new results for Tschebyscheff approximation by these splines with free knots are sketched, the closure of $S_{m,k}$ is characterized for classes of stiff regular splines steep at the boundary, and an upper bound for the number of zeros of $s-s^*$, where $s, s^* \in S_{m,k}$ is quoted leading to the length of a (sufficient) alternant.

1 Introduction

For practical purposes it is useful to allow for approximating functions which share certain properties with the function being approximated thus motivating the definition of regular splines which sets up the spline from portions of (mildly) nonlinear families.

Let $I = [\alpha,\beta]$ and consider subdivision into m non-zero intervals $I_j = [x_{j-1}, x_j]$, $x_o = \alpha$, $x_m = \beta$. Hence we have m-1 interior points, called knots $x_1 < x_2 < \ldots < x_{m-1}$. Let k specify the degree of smoothness, $k \geq 2$.

Given are m classes $\mathcal{T}_j$ of functions

$$t_j(x; x_{j-1}, x_j, c, d) \in C^k(x_{j-1}, x_j)$$

continuously depending, as are their derivatives, on the four parameters x_{j-1}, x_j, c, d, where c and d vary in appropriate open intervals. Sometimes higher differentiability properties of t_j are needed. (Note that the intervals I_j may vary.) Then

$$S_{m,k} := \{u \mid u \in C^k(I),\ \exists\, I_1,\ldots,I_m,\ I_j \subset I \text{ such that } u|_{I_j} = p_j + t_j,\ p_j \in \mathcal{P}_{k-1},\ t_j \in \mathcal{T}_j,\ j=1,\ldots,m\}$$

describes the class of regular splines with free knots. $\mathcal{P}_{k-1}$

denotes the polynomials of degree k-1 or less.

The class $\mathcal{T}$ should be regular, as first introduced by Schaback [13], i.e. it is possible to parametrize the class by means of the values of $t^{(k)}$ at two different points, in particular x_{j-1} and x_j. In other words, if t^*, $t \in \mathcal{T}$ and

$$D^k(t^*-t)(z) = o \quad \text{for} \quad z = x_{j-1} \quad \text{and} \quad x_j,$$

then $t^*(x) \equiv t(x)$ in $[x_{j-1}, x_j]$.
Equivalently the differences of the kth order derivatives of two functions of $\mathcal{T}$ can have at most one separated zero.

Typical examples of $\mathcal{T}$ are

$$t^o = (d\cdot x+c)\cdot x^k, \quad t^1 = \frac{c}{d-x}, \quad t^2 = c\cdot\exp(dx), \quad t^3 = c\,\log(x+d),$$

with obvious restrictions on c, d depending on the domain of x. If c and d are arbitrary reals, t^o just produces the polynomial splines, but with other restrictions other results will arise.

2 Survey on Interpolation

The above properties suffice to deal with the interpolation problem, data prescribed in the way familiar from polynomial splines. For the class t^1 we obtain the so called special rational splines in $S_{m,k}$. For $k = 2$ Schaback [12] treated the problem extensively and established uniqueness. To obtain existence he rewrote the interpolation equations as variational equations of an optimization problem. Recently Baumeister [4] dualized the optimization method to get a new definition of certain classes of regular splines, compare also Baumeister, Schumaker [5] , Schumaker [15]. For sufficiently fine subdivisions of I the interpolation with regular splines is closely related to the polynomial spline interpolation. It can be treated by a perturbation method as I showed in Werner [18] for k = 2; the general case was solved by Arndt [1].

To conclude this survey it should be added that regular splines are a powerful tool in handling initial value problems of differential equations, compare Runge [11], Werner [18], [19] and boundary value problems, compare the forthcoming Brunel thesis of Ph. Moore [10]. In this case the freedom of the localisation of the knots is particularly valuable although with proper choice of $\mathfrak{T}$ one may do quite well with evenly spaced knots.

3 Tschebyscheff Approximation with free knots

To develop the theory of T-Approximation we will introduce two additional assumptions upon the class $\mathfrak{T}$. The function $t \in \mathfrak{T}$ shall be stiff, i.e. $D^k t > o$, respectively $D^k t < o$. This excludes heavily oscillating functions and is useful in considering sequences in $S_{m,k}$.

While boundedness of the function t in an interval guaranties boundedness of the parameters in case of polynomials, i.e. for example in t^o, this may be quite different in other cases. To get a realistic picture of these circumstances we analyze the special rational splines more closely.

Consider the one parameter family

$$\tilde{t}(z,d) := c_k \cdot \frac{z^k}{d-z} = c_k \cdot \frac{d^k}{d-z} - p(z,d),$$

where $p(z,d) = (z^{k-1}+z^{k-2}\cdot d+\ldots+d^{k-1})\cdot c_k$, $c_k = \frac{1}{(k-1)!}$

in $z \in [-1,o]$, say, and $o < {}^1/_d + 1 < \infty$.

The following limit values for $d \to o$ are obvious:

	$D^k\tilde{t}$	$D^j\tilde{t}$ $(j < k)$	
$z = o$	∞	o	(*)
$z < o$	o	$-D^j p(z,o)$	

Observe the jump of the (k-1)st derivatives at z = o. The

analogous situation can be found at the left endpoint of the interval.

We say the class $\mathcal{T}$ is <u>steep</u> at the boundary if there is a (k-1)st degree polynomial such that

$$\tilde{t}(x,d) = t(x;\, x_{j-1},\, x_j,\, 1,\, d) - p(x;\, x_{j-1},\, x_j,\, d)$$

has the limits given in (*) if $z = x - x_j$ for $d \to o$ and if the analog is true with $z = x - x_{j-1}$, for $d \to \infty$, say. The parameter c should enter multiplicatively for the class.

To investigate existence of best approximations we consider the <u>closure of $S_{m,k}$</u> under compact convergence. One can even show <u>the compactness</u> of bounded sequences of $S_{m,k}$.

If $J \subset (\alpha,\beta)$, the uniform boundedness of any sequence $\{s^{(v)}\}$ in I implies uniform boundedness of all derivatives up to order k-1 in J due to the stiffness (which implies monotonicity of $D^{k-1}s^{(v)}$ and boundedness of appropriate difference quotients. By taking subsequences one may assume that a sequence, again denoted by $s^{(v)}$, has been found such that $D^j s^{(v)} \to D^j s$ uniformly in J for $j = o,\ldots,k-2$; furthermore, that $\lim x_j^{(v)} = x_j$ exist, i.e. the knots converge, as do the parameters $d_j^v \to d$ and $c_j^v \to c_j$ for $v \to \infty$, where the limit may be ∞ .
It is easily seen that in any subinterval of J free of limit knots x_j, the limit function s is either equal p_j, a polynomial of degree (k-1) ("polynomial"), or $p_j + t_j$ ("regular"). So it remains to investigate the behaviour in a neighbourhood U of an interior knot x_j. If $s \in C^{i-1}(U) \setminus C^i(U)$ the knot is called of the <u>ith kind</u>. A knot is <u>n-fold</u>, if n limit knots $x_j,\ldots,x_{j+n-1}$ coincide. 1-fold knots are called simple.

1) At simple knots we observe the same phenomena as for fixed knot approximation, compare Werner [16,17]. If at x_j locally the function s is:

   i) regular at both sides of x_j, then $s \in C^k(U)$,

ii) regular at one, polynomial at the other side, x_j is of kth kind,

iii) polynomial at both sides, x_j is of (k-1)st kind.

2) At two fold knots we may think of one subinterval shrinking to zero length and in this interval s becoming polynomial, this allows an extra discontinuity and leads to the following additional new cases. If at $x_j = x_{j-1}$ the function s is locally:

iv) as in i) above, then x_j is of kth kind,

v) as in ii) above, then x_j is of (k-1)st kind.

3) Finally, if the knot x_j is three-fold (or even more) then we may think that two polynomial intervals of zero length may appear. This makes it possible that s is regular at both sides of x_j, but in spite of this, x_j is of (k-1)st kind.

Since there cannot exist (k-2)nd kind knots due to the continuity of $D^{k-2}s$, no other cases are possible. To complete the description of $\bar{S}_{m,k}$, one has to add that interior polynomial sections always are the limit of at least two regular sections of the converging sequence; i.e. it seems that every polynomial interval absorbs one knot.

If we define the sum of interior knots (counting multiplicity as 1,2, or 3) plus the number of interior polynomial sections of s as its order ord(s), one has to work out that $s \in \bar{S}_{m,k} \setminus S_{m,k}$ of order $k' \leq k$ is the limit of functions of $S_{m,k'}$ under compact convergence.

These statements are generalizations of the results for special rational splines obtained by Schomberg [14]. It is now a straightforward conclusion to establish existence of best approximations s^* in $\bar{S}_{m,k}$, but a quite different matter to find out when s^* will lie in $S_{m,k}$.

To establish sufficiency criteria for best approximations (as length of alternants) in the standard way, one may count zeros of differences $r = s-\bar{s}$, where $s,\bar{s} \in \bar{S}_{m,k}$. The considerations of Schomberg [14] immediately carry over to show that this number (counting multiplicities) has the upper bound

$$\mathrm{ord}(s) + \mathrm{ord}(\bar{s}) + k+1 - \# \text{ (common knots of } s \text{ and } s').$$

The details of these results will be contained in a forthcoming report.

References

1 Arndt, H., Interpolation mit regulären Spline-Funktionen, Dissertation Münster, 1974.

2 Barrar, B.R. and H.L. Loeb, Existence of Best Spline Approximations with Free Knots, J. Math. Anal. & Appl., 31 (1970), 383-39o.

3 Barrar, B.R. and H.L. Loeb, Spline Functions with Free Knots as the Limit of Varisolvent Families, J. Approximation Theory, 12 (1974), 70-77.

4 Baumeister, J., Extremaleigenschaften nichtlinearer Splines, Dissertation München, 1974.

5 Baumeister, J. and L.L. Schumaker, Nonlinear Classes of Splines and Variational Problems, J. Approximation Theory.

6 Braess, D., Chebyshev Approximation by Spline Functions with Free Knots, Numer. Math., 17 (1971), 357-366.

7 Braess, D., On the Degree of Approximation by Spline Functions with Free Knots, Aequationes Math., 12 (1975), 80-81.

8 Braess, D. and H. Werner, Tschebyscheff-Approximation mit einer Klasse rationaler Splinefunktionen II, J.Approximation Theory 10 (1974), 379-399.

9 Loscalzo, F.R. and T.D. Talbot, Spline Function Approximation for Solutions of Ordinary Differential Equations, SIAM J.Numer.Anal., 4 (1967), 433-445.

10 Moore, P., Regular Spline Solutions to Two Point Boundary Value Problems. Ph. D. Thesis, Brunel University, 1976.

11 Runge, R., Lösung von Anfangswertproblemen mit Hilfe nichtlinearer Klassen von Spline-Funktionen, Dissertation Münster, 1972.

12 Schaback, R., Spezielle rationale Splinefunktionen, J. Approximation Theory 7 (1973), 281-292.

13 Schaback, R., Interpolationen mit nichtlinearen Klassen von Splinefunktionen, J. Approximation Theory 8 (1973), 173-188.

14 Schomberg, H., Tschebyscheff-Approximation durch rationale Splinefunktionen mit freien Knoten, Dissertation Münster, 1973.

15 Schumaker, L.L., Toward a Constructive Theory of Generalized Spline Functions, in Spline Functions Karlsruhe 1975, Lecture Notes 501, Springer (1976), 265-331.

16 Werner, H., Tschebyscheff-Approximation mit einer Klasse rationaler Splinefunktionen, J. Approximation Theory 10 (1974), 74-92.

17 Werner, H., Tschebyscheff-Approximation nichtlinearer Splinefunktionen, in K.Böhmer-G.Meinardus-W.Schempp: Spline-Funktionen, BI-Verlag Mannheim-Wien-Zürich 1974, 3o3-313.

18 Werner, H., Interpolation and Integration of Initial Value Problems of Ordinary Differential Equations by Regular Splines, SIAM J. Numer. Anal. 12 (1975), 255-271.

19 Werner, H., An Introduction to Regular Splines and Their Application for Initial Value Problems of Ordinary Differential Equations. TR/53 (1975), Brunel University.

20 Werner, H. and R. Schaback, Praktische Mathematik II, Springer-Verlag, Berlin-Heidelberg-New York 1972.

H. Werner
Rechenzentrum
Universität Münster
Roxeler Str. 60
44 Münster
Germany

AN APPROXIMATION PROOF IN THE THEORY OF COMPLEX L_1-PREDUALS

Daniel Wulbert

Abstract. A short proof is given for the Hirsberg-Lazar characterization of complex L_1-predual subspaces of $C(X, \mathbb{C})$. The proof depends on the approximation of L_1-preduals by certain finite dimensional subspaces, and uses a Korovkin type approximation argument about the convergence of contractive projections.

I Introduction

The main results related to this lecture are about the structure of complex L_1-predual Banach spaces. However a preliminary part of the development is a short proof of the Hirsberg-Lazar characterization of such subspaces of $C(X, \mathbb{C})$. The proof depends on the approximation of L_1-preduals by certain finite dimensional subspaces. The Hirsberg-Lazar proof uses the simplex structure that E.G. Effros showed was associated with a complex L_1-predual. This talk will present the approximation theoretic approach.

The result by Hirsberg and Lazar is the following:

THEOREM 1.1 *If* E *is a subspace of* $C(X, \mathbb{C})$ *which contains the constants, the following are equivalent*:

(a) E *is an* L_1-*predual*

(b) E *is selfadjoint and* $\operatorname{Re}E$ *is a real* L_1-*predual*.

II Preliminaries and Notation

The following two theorems characterize L_1-preduals by their property of being approximatable by subspaces of the type

ℓ_∞^n. Both results apply to real and complex Banach spaces.

THEOREM 2.1 *If* E *is a separable* L_1*-predual Banach space contained in a Banach space* F *then there is a sequence of norm one projections* $\{P_i\}_{i=1}^\infty$ *defined on* F *such that*

1. $P_i f \in E$ *for all* $f \in F$ *and*
2. $P_i g \to g$ *in norm for all* $g \in E$.

THEOREM 2.2 E *is an* L_1*-predual if and only if for each* $\varepsilon > 0$ *and each finite set* $S \subseteq E$, *there is a finite dimensional subspace* $F \subseteq E$ *such that*

1. F *is isometric with* ℓ_∞^n *for some* n *and*
2. $\sup\{\inf\{\|s - f\| : f \in F\} : s \in S\} < \varepsilon$.

Both results are known; and, for example, either can be found in [Lazar and Lindenstrauss, 1966] or as variants of results there.

We will use the rest of the section to record our notation. We will use $\mathbb{R}$ and $\mathbb{C}$ to denote the real and complex fields. For a compact Hausdorff space X, $C(X, \mathbb{C})$ represents the continuous complex valued functions on X, equipped with supremum norm, and $\ell_\infty^n(\mathbb{C})$ is equivalent to $C(Y, \mathbb{C})$ for a set Y containing n points. For a measure space (X,Σ,μ), $L_1(X,\Sigma,\mu,\mathbb{C})$ consists of the integrable complex valued functions on (X,Σ,μ). Analogous definitions hold for the real Banach spaces obtained when replacing $\mathbb{C}$ by $\mathbb{R}$ above.

For a Banach space E, E^* and $S(E)$ represent the dual of E and the unit ball of E respectively. The extreme points of a convex set K are the set $\mathrm{ext}(K)$. A Banach space E is an L_1-predual if there is a measure space (X,Σ,μ) such that E^* is isometric with $L_1(X,\Sigma,\mu)$.

III Proof of Theorem 1.1

PROPOSITION 3-1. *Let* E *be a subspace of* $C(X, \mathbb{C})$ *that contains the constants. If* E *is an* L_1*-predual then* E *is selfadjoint.*

Proof. Although this is not obvious, it is known, that we may assume that E is separable and that E separates the points of X.

Let

$$(3.1) \quad D = \text{closure}[\text{span } \mathrm{Re}E] \subseteq C(X, \mathbb{C}).$$

Then

$$(3.2) \quad \partial E = \partial \mathrm{Re}E = \partial \mathrm{Re}D = \partial D,$$

where ∂ denotes the Choquet boundary of a space.

Let P_i be the projections of theorem 2.1. Then

$$(3.3) \quad (P_i d)(x) \to d(x) \quad \text{for} \quad x \in \partial E \quad \text{and} \quad d \in D.$$

From Choquet theory and the Lebesgue dominated convergence theorem, this implies that $P_i d \to d$ weakly (since $\partial E = \partial D$). Since $P_i d \in E$ we conclude that $D \subseteq E$ and E is selfadjoint.

PROPOSITION 3-2 *Let* E *be a selfadjoint subspace of* $C(X, \mathbb{C})$. *Then* E *is a complex* L_1*-predual if and only if* ReE *is a real* L_1*-predual*.

Proof. The proof uses theorem 2.2. First assume E is a complex L_1-predual. Let G be a complex linear subspace of $C(X, \mathbb{C})$ that is isometric to $\ell_\infty^n(\mathbb{C})$, and suppose f is a real function such that $d(f, G) < \varepsilon$. We will show there is a real linear subspace S of $\mathrm{Re}G \subseteq E$ that is isometric with $\ell_\infty^n(\mathbb{R})$ and for which $d(f, S) < 2\varepsilon$. By theorem 2.2, this will prove the result. Since G is isometric with ℓ_∞^n there exists

a basis $\{g_i\}$ of G and points $\{x_i\}_{i=1}^n \subseteq x$ such that

(a) $g_i(x_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

and

(b) for each $y \varepsilon X$ there are scalars λ_i such that

both

(i) $\Sigma|\lambda_i| \leq 1$ and

(ii) $\Sigma\alpha_i g_i(y) = \Sigma\lambda_i\alpha_i$ for any set f scalars α_i.

Now let

(3.4) $S = \text{real span}\{\text{Re}g_i\}_{i=1}^n \subseteq E$.

If follows that the map

(3.5) $f \to (f(x_1), f(x_2), \ldots, f(x_n))$

is an isometry of S onto $\ell_\infty^n(\mathbb{R})$. For if $x \varepsilon X$ and $\alpha_i \varepsilon \mathbb{R}$,

(3.6) $|\Sigma\alpha_j \text{Re}g_j(x)| \leq |\Sigma\alpha_j g_j(x)| \leq |\Sigma\lambda_i\alpha_j| \leq \max|\alpha_j|$.

Now to verify the estimate $d(f, S) < 2\varepsilon$, let $\alpha_j \varepsilon \mathbb{C}$ be such that $\|f - \Sigma\alpha_j g_j\| < \varepsilon$.
One verifies that

(3.7) $|\text{Im}\alpha_j| \leq \varepsilon$, and $\|\Sigma\alpha_j g_j - \Sigma(\text{Re}\alpha_j)g_j\| < \varepsilon$.

Thus if

(3.8) $q = \text{Re}\Sigma(\text{Re}\alpha_j)g_j$,

then

(3.9) $\|\Sigma \mathrm{Re}(\alpha_j g_j) - q\| < \varepsilon$.

But since $\|\Sigma \mathrm{Re}\alpha_j g_j - f\| < \varepsilon$, $\|q - f\| < 2\varepsilon$.

This proves the case in which E is a complex L_1-predual. In fact this shows that ReE is a real L_1-predual even when E is not selfadjoint.

If ReE is a real L_1-predual one shows that the complex span, of an isomotric copy of $\ell^n_\infty(R) \subseteq$ ReE is isometric with $\ell^n_\infty(\mathbb{C})$. The proof is almost exactly as above but shorter since it dosen't require computing the estimate above. This part is deleted.

References

Effros, E. G. Structure in simplexes, Acta Math., 117 (1967), 103-121.

Effros, E. G. On a class of complex Banach spaces, Illinois J. Math., 18 (1974), 48-59.

Hirsberg, B., and A. J. Lazar. Complex Lindenstrauss spaces with extreme points, Trans. Amer. Math. Soc., 186 (1973), 141-151.

Lacey, E. The Isometric Theory of Classical Banach Spaces, Springer-Verlag, New York, 1974.

Lazar, A. J., and J. Lindenstrauss. On Banach spaces whose duals are L_1 spaces, Israel J. Math., 4 (1966), 205-207.

Lazar, A. J. The unit ball in conjugate L_1 spaces, Duke Math. J., 39 (1972), 1-8.

Lazar, A. J., and J. Lindenstrauss, Banach spaces whose duals are L_1 spaces and their representing matrices, Acta Math., (1971), 165-194.

Lindenstrauss, J. Extensions of compact operators, Memoirs Amer. Math. Soc., 48 (1964).

Wulbert, D. E. Convergence of operators and Korovkin's theorem, J. Approx. Theory, 1 (1968), 381-390.

Wulbert, D. E. Real Structure in Complex L_1-Preduals, preprint 1976.

Daniel E. Wulbert
University of California, San Diego
La Jolla, California 92093
U.S.A.

The research was supported by the National Science Foundation.

APPROXIMATION IN A NORMED SPACE

New foundations and methods

Marc Zamansky

In this paper we outline a novel approach to obtaining convergence, approximation, and saturation results for certain classes of linear approximation processes. A more detailed (preliminary) exposition of these ideas has appeared as a preprint: Zamansky, Approximation, XI + 84 p., 1975, Institut Henri Poincare, Paris. Here we have changed the order and stated some results (cf. (4.1), (4.6), (4.7)) in more general form.

The origin of this theory is a very simple proof of the theorem about the saturation class of the first arithmetic mean (1950). It seems that only Favard in 1957 and Aljancic in 1958 have used this idea, but in special cases. Here is the idea: if S_n is a sequence of numbers, $T_n = (S_o + \dots + S_{n-1})/n$ and $0 < \alpha < 1$, then

$$S_n - s = 0(1)/n^{\alpha} \Longleftrightarrow T_{n+1} - T_n = 0(1)/n^{\alpha+1}.$$

We begin in section 1 by defining the <u>classes of approximations</u>: they are all sequences $\rho = (\rho_n)$, $\rho > 0$, $\rho \downarrow$ that satisfy some conditions. The <u>classes of</u> (linear) <u>processes</u> of approximation are defined by natural conditions in reference to processes $[1 - t^p]$ (typical means) which are "<u>standard-processes</u>."

We prove that, under general conditions, any "usual" process is equivalent to a standard--process for a well defined approximation class.

For example, we apply this theorem and the intermediate theorems to periodic functions and obtain in a few pages all fundamental results, known or new. It is superfluous to distinguish convergence, approximation, or saturation. We even obtain the best asymptotic constants in general cases. We note that Fourier transforms do not play any role.

1 <u>Classes of approximations</u>

An approximation is a real, positive function $\rho \downarrow$ of $\theta \geq \theta_o$ and a class of approximations is the set of ρ which

satisfy some conditions. We define the class $\{\Lambda\}$ by

(1.1) $\rho_{\lambda n} = 0(1)\rho_n$, $0 < \lambda < 1 \Longleftrightarrow \sum_{\lambda n \le k \le n} \rho_k = 0(1)\, n\, \rho_n$

(where $0(1)$ is dependent on λ, but is in n). $\{\Lambda\}$ contains all usual approximations but not exponential approximations.

(1.2) Let $A_n = a_o + \ldots + a_n \neq 0$ and $\sigma_n(a,\rho) = (\sum_0^n |a_k|\rho_k)/|A_n|$, $\tau_n(a,\rho) = \sum_n^\infty (|a_k|/|A_{k-1}|)\rho_k$ which define the functions $\sigma(a):\rho \to \sigma(a,\rho)$, $\tau(a):\rho \to \tau(a,\ )$. We write $\sigma(\alpha)$ and τ when $a_k = (k+1)^\alpha$, $\alpha > -1$.

The classes $\{\sigma(\alpha)\}$, $\{\tau\}$ are defined by $\sigma(\alpha,\rho) = 0(1)\rho$, $\tau(\rho) = 0(1)\rho$ (it follows that $\sigma(\alpha,\rho) \sim \rho$ and, if $p \in \{\Lambda\}$, $\tau(\rho) \sim \rho$).

(1.3) We establish the technical properties that will play a role. For example: $\{\sigma(\alpha)\} \subset \{\Lambda\}$, $(\sigma_n(\alpha,\rho)) \in \{\Lambda\}$, $\sigma(\beta) \circ \sigma(\alpha) = 0(1)\ \sigma(\inf\ (\alpha,\beta))$, $\sigma(\alpha) \circ \tau$ and $\tau \circ \sigma(\alpha)$ are $0(1)\ (\sigma(\alpha) + \tau)$, etc...

(1.4) Because $\sigma(\alpha,\rho) = 0(1)\rho \Rightarrow \sigma(\alpha,\ o(\rho)) = o(1)\rho$ and, if $\tau_n(\rho) = 0(1)$, $\tau_n(o(\rho)) = o(1)\tau_n(\rho)$, all results will be true when we take $o(\rho)$ instead of ρ.

(1.5) If $\varepsilon_n > 0$, $\varepsilon_n \to 0$, there exists $\rho_n \downarrow 0$, $\rho_n \geq \varepsilon_n$ and $\rho \in \{\sigma(\alpha)\}$ for all $\alpha > -1$.

A consequence is that all theorems of approximation will be true if we consider only the convergence.

2 Classes of processes

Any (linear) process for a series Σu_k in a normed space E gives $\sum_k c_{n,k} u_k$; therefore we define a process by a sequence (ϕ_n) of functions of $t = k/n$ ($n = 1, 2, \ldots, k = 0, 1, 2, \ldots$) such that $\phi_n(t) = c_{n,k}$.

In the interest of clarity, we distinguish the processes $[\phi_n]$ or $[\phi_n(t)]$, where $\phi_n(t) = 0$ if $t < 0$, $t > 1$, and the processes $[\phi_n[$ or $[\phi_n(t)[$, and where $\phi_n(t) = 0$ if $t < 0$. We write $T_n(\phi_n, u_k) = \sum_k \phi_n(k/n)u_k$.

Let $\Delta_1(\phi_n) = \phi_n(t) - \phi_n(t + 1/n)$, $\Delta_2(\phi_n) = \Delta_1(\Delta_1(\phi_n))$,

(2.1) The class of processes $\{[p, a_n, \varepsilon]\}$ is the set of $[\phi_n]$ such that the ϕ_n are uniformly bounded and for a finite number of ξ_i with $0 < \eta' < \eta < \xi_i < \xi_{i+1} \leq 1$,

(1) in $[\eta', 1]$, $n\Delta_1(\phi_n) = 0(1)$ uniformly with respect to t and n;

(2) in $[\eta', \xi_1[, \ldots,]\xi_i, \xi_{i+1}[$, $n^2\Delta_2(\phi_n) = 0(1)$ likewise

(3) in $[0, \eta]$, $\phi_n(t) = \phi_n(0) - t^p(a_n + \psi_n(t))$ where $p > 0$, a_n = constant and for a $\varepsilon \geq 0$ and $0 < t \leq \eta$: $t^{-\varepsilon}\psi_n(t) = 0(1)$, $n\, t^{1-\varepsilon}\Delta_1(\psi_n) = 0(1)$, $n^2\, t^{2-\varepsilon}\Delta_2(\psi_n) = 0(1)$ uniformly with respect to t and n.

(2.2) The class $\{[p, a_n, \varepsilon, \beta[\}$ is the class of the processes $[\phi_n[$ which satisfy conditions (2.1) and

(4) for $\eta' \leq t$ and a $\beta \geq 0$, $t\phi_n(t) = 0(1)$, $n\, t^{\beta+1}\Delta_1(\phi_n) = 0(1)$, $n^2 t^{\beta+2}\Delta_2(\phi_n) = 0(1)$ uniformly.

(2.3) The standard-processes are $[1 - t^p]$ and we will answer the question: if a standard-process gives an approximation ρ, what is the approximation for $[\phi_n]$ or $[\phi_n[$, and inversely?

N.B.--If we compare $[\phi_n]$ with $[1]$, the hypotheses which concern $\Delta_2(\phi_n)$, are useless.

(2.4) With some precautions, but without new hypotheses, the results are true for the processes $\phi_n(k/n^{1/p})$ or more generally for $\phi_x(k/x)$.

3 General Theorems in a normed space

We introduce the notation $([\phi_n]; \rho_n) := T_n([\phi_n], u_k) - \phi_n(0)s - \phi_n(1)(T_n([1], u_k) - s) = 0(1)\rho_n$, $s \in E$ and $([\phi_n[; \rho_n) := T_n([\phi_n[, u_k) - \phi_n(0)s = 0(1)\rho_n$.

(3.1) THEOREM A. We have

(3.1.1) $\rho \in \{\Lambda\}$, $([1]; \rho_n) \Rightarrow ([\phi_n]; (|a_n| + 1)\sigma_n(p - 1, \rho))$

(3.1.2) $\rho \in \{\Lambda\}$, $([1 - t^p]; \rho_n) \Rightarrow ([\phi_n]; (|a_n| + 1)\sigma_n(p + \varepsilon - 1, \sigma(p, \rho)))$.

(3.1.3) If $\sum_k \phi_n(k/n)u_k$ converges and $\rho \in \{\Lambda\}$, one has:

$\beta > 0$, $([1 - t^p]; \rho_n) \Rightarrow ([\phi_n[; (|a_n| + 1)\sigma_n(p + \varepsilon - 1, \sigma(p, \rho)) + \sigma_n(p, \rho))$

$\beta = 0$, $([1 - t^p]; \rho_n) \Rightarrow ([\phi_n[; (|a_n| + 1)\sigma_n(p + \varepsilon - 1, \sigma(p, \rho)) + \tau_n(\sigma(p, \rho))$.

(3.1.4) For $\phi_n(k/n^{1/p})$ there are analogous results.

(3.2) THEOREM B. (Principal Part) Let $p > 0$, $q + 1 > 0$, $q + 1 \neq p$. Then

(1) If E is normed, $([1 - t^p]; \rho_n) \Rightarrow ([t^p - t^{q+1}]; \sigma_n(q, \rho))$ when $q < p$, or $q > p$ and $\sigma(p) = 0(1)\sigma(q)$, or $q = p$ and $\sigma(p) \in \{\sigma(p)\}$.

(2) If E is a Banach space, $\tau_n(\rho) = 0(1)$ and $([t^p - t^{q+1}]; \rho_n) \Rightarrow \exists\, s \in E$ and $([1 - t^p]; \tau_n(\rho))$ when $p \neq q$ and $\rho \in \{\sigma(\inf(p, q))\}$ or $\rho \in \{\sigma(p) \circ \sigma(p)\}$.

(3.3) Examples.

(3.3.1) Theorem A contains many know results about convergence, but here we have stated only results about approximations. If we consider, for example, $[1 - t^2] \circ [1 - t]$, we obtain the Cesaro processes. The Riemann process $\phi(t) = \sin^p t/t^p$ is a very good example for 3.1.3 ($\beta = 0$ if $p = 2$, $\beta > 0$ if $p > 2$). For 3.1.4 we mention Gauss-Weierstrass, de la Vallee-Poussin, generalized Abel processes and many others.

(3.3.2) The processes $[\phi]$ (cf. the work of Sz. Nagy in 1948 but only for the periodic functions in the space C) are a special case.

(3.3.3) It is possible to write the results with ρ = constant and $s = 0$. We obtain an upper bound for $\|T_n(\phi_n, u_k)\|$ (cf. theorem N in 4.1).

4 Approximation in a normed space possessing a biorthogonal system

We suppose that the sequence $e_i \varepsilon E$ and sequence (c_j) of linear forms compose a biorthogonal system. For $f \varepsilon E$, we consider Σu_k where $u_k = c_k(f)e_k$. We write $T_n(\phi_n, f)$ instead of $T_n(\phi_n, u_k)$ and $P = \sum_o^n \alpha_k e_k$, $P^{[m]} = \sum_o^n k^m \alpha_k{}_k$, $m > 0$.

(4.1) THEOREM N. <u>If the operators</u> $T_n([1 - t])$ <u>(the first arithmetic means) are continuous and</u> $\|T_n([1 - t])\| = 0(1)$, <u>then for</u> $[\phi_n] \varepsilon \{[p, a_n, \varepsilon]\}$, $\phi_n(1) = 0$, $a_n = 0(1)$, <u>one has</u> $T_n([\phi_n]) = 0(1)$.

(4.2) THEOREM C. <u>If</u> $P_n - f = 0(\rho_n)$ <u>and</u> $\|P_n^{[m]}\| = 0(1)n^m\|P_n\|$, <u>then</u>

$$(1) \quad P^{[m]}_{a^{k+1}} - P^{[m]}_{a^{k_o}+1} = 0(1) \sum_{\nu=a^{k_o}}^{\nu=a^k-1} \nu^{m-1} \rho_\nu, \text{ a integer} \geq 2.$$

$$(2) \quad P^{[m]}_{n+1} - P^{[m]}_{n_o+1} = 0(1)\left((n_o + 1)^m \rho_{n_o} + \sum_{\nu=n_o}^{\nu=n} (\nu + 1)^{m-1} \rho_\nu\right)$$

$$(3) \quad P_n^{[m]} = 0(1)n^m\sigma_n(m - 1, \rho) \text{ (first proof in 1949).}$$

(4.3) $P_n - f = 0(\rho_n) \Rightarrow T_n([1 - t^p], f) - f = 0(1)\sigma_n(p - 1, \rho)$.

(4.4) <u>If</u> $\phi_n(0) = 1$, $\phi_n(1) = 0$, $[\phi_n] \varepsilon \{[p, a_n, \varepsilon]\}$, <u>then</u> $T_n([\phi_n], f) - f = 0(\rho_n) \Rightarrow a_n(T_n([1 - t^p], f) - f) = 0(1)\sigma_n(p - 1 + \varepsilon, \rho)$.

(4.5) <u>We suppose that</u> E <u>is a Banach space,</u> $[\phi_n] \varepsilon$ $\{[p, a_n, \varepsilon, \beta]\}$, $\phi_n(0) = 1$, $0 < \inf |a_n| \leq \sup |a_n| < +\infty$. <u>Moreover, we suppose for a sufficiently small</u> $\eta > 0$ <u>and</u> $0 < \eta \leq t \leq 1$, <u>one has</u> $\inf_{n,t} |1 - \phi_n(t)| > 0$. <u>Finally, we suppose that, if</u> $\varepsilon = 0$, <u>for</u> $0 < t \leq \eta$ <u>one has</u> $\inf_{n,t}(|1 - \phi_n(t)|/t^p) > 0$. <u>Then, if</u> $\rho \; \varepsilon \; \{\sigma(p)\}$ <u>and</u> $\tau_n(\rho) = 0(1)$,

$$T_n([\phi_n[, f) - f = 0(\rho_n) \Rightarrow T_n([1 - t^p], f) - f = 0(1)\tau_n(\rho)$$

(4.6) <u>Equivalence between</u> $[\phi_n]$ <u>or</u> $[\phi_n[$ <u>and</u> $[1 - t^p]$. Let E be a normed linear space possess a biorthogonal system. We suppose that the operators $T_n([1 - t])$ are continuous and $\|T_n([1 - t])\| = 0(1)$. We suppose $0 < \inf |a_n| \leq \sup |a_n| < +\infty$ and we consider the processes which satisfy (4.4) or (4.5). Then $[\phi_n] \sim [1 - t^p]$ for all $\rho \; \varepsilon \; \{\sigma(\inf(p + \varepsilon - 1, p))\}$; and if E is a Banach space $[\phi_n[\sim [1 - t^p]$ for all $\rho \; \varepsilon \; \{\sigma(\inf(p + \varepsilon - 1, p))\} \cap \{\tau\}$

N.B. One may extend these results to other processes, for example $\phi_n(k/n^{1/p})$

(4.7) <u>Equivalence between</u> $[\phi_n]$ <u>and</u> $[1]$. <u>If</u> $0 < \inf|\phi_n(1)| \leq \sup|\phi_n(1)| < +\infty$ <u>and</u> $\rho \; \varepsilon \; \{\sigma(p - 1)\}$, <u>then</u> $[\phi_n] \sim [1]$ <u>for the approximations</u> $(|a_n| + 1)\rho_n$.

(4.8) <u>Saturation</u>. <u>Saturation takes place if the</u> c_j <u>are continuous and we obtain the saturation classes</u>.

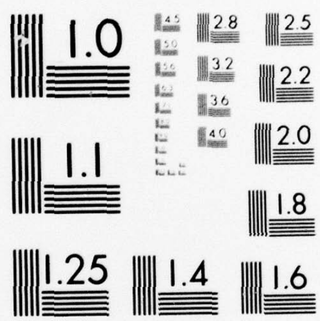
5 <u>Example: approximation of periodic functions</u>

We give only one example. It is unnecessary to specify which usual topology is used. With the previous results, we find immediately the differential properties from ρ_n to ω_m, (m-th modulus of continuity, m integer ≥ 1). If we want to

AD-A033 032   TEXAS UNIV AT AUSTIN DEPT OF MATHEMATICS   F/G 12/1
APPROXIMATION THEORY II
NOV 76   LORENTZ, G.; CHUI, C.; SCHUMAKER, L.
UNCLASSIFIED   AFOSR TR-76-1242   N/L

7 OF 7
AD
A033032

END
DATE
FILMED
9 -78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

prove the inverse properties, it is sufficient to prove only the following theorems.

(5.1) THEOREM J. For any integer $m \geq 1$, there is a sequence of polynomials P_n of degree n, such that $P_n - f = 0(1)\omega_m(1/n)$. (We give a new proof).

(5.2) Approximation by $[1 - t^p]$.

If $0 < p \leq m$, $T_n([1 - t^p], f) - f = 0(1)\sigma_n(p - 1, \omega_m)$

If $p > m$, $T_n([1 - t^p], f) - f = 0(1)\omega_m(1/n)$

If $p = 2m$, $T_n([1 - t^{2m}], f) - f = 0(1)\omega_{2m}(1/n)$

(5.3) If the conjugate $\tilde{f} \in E$ and $\tau_n(\rho) = 0(1)$, then (cf. 4.3)

$$P_n - f = 0(\rho_n) \Rightarrow T_n([1 - t^p], \tilde{f}) - \tilde{f} = 0(1)(\sigma_n(p - 1, \rho)$$

$$+ \tau_n(\rho)).$$

We obtain quickly all relations between derivatives and approximations, between the ω_m of f and $\tilde{f}$, between the ω_m and the best approximations, and the saturation classes of processes $[\phi_n]$ or $[\phi_n[$.

N.B.--For the periodic functions, it is possible to give best hypotheses for Theorem 4.6: for $[\phi_n]$ (resp. $[\phi_n[$) in L^r, $1 < r < +\infty$, $\phi_n(1) = 0(1)$ (resp. $\phi_n(t) = 0(1)$) and in the other spaces, $\phi_n(1) = 0(1/\log n)$ (resp. $\phi(t) = (1/\log t)$, $t \to +\infty$).

(5.4) If $\varepsilon > 0$, the results give the approximations:

for $[\phi_n]$, "from convergence to saturation inclusively";

for $[\phi_n[$, "from approximations such that $\tau_n(\rho) = 0(1)$ to saturation inclusively"

(5.5) <u>Best asymptotic constants</u>. (Example).

THEOREM. <u>Let</u> $[\phi_n] \varepsilon \{[p, a_n, \varepsilon]\}$, (resp. $[\phi_n[\varepsilon \{[p, a_n, \varepsilon, \beta]\}$ <u>where</u> $a_n \neq 0$, $a_n = O(1)$, $\varepsilon > 0$, $\beta > 0$. <u>Let</u> $2m$ <u>be an integer</u> ≥ 2 (resp. $2m + 1 \geq 1$).

<u>If</u> $f^{(2m)} \varepsilon E$ <u>and</u> $p = 2m$, <u>then</u>

$$T_n(\phi_n, f) - f + \frac{(-1)^m}{n^{2m}} a_n f^{2m} = o\left(\frac{1}{2m}\right)$$

<u>If</u> $\tilde{f}^{(2m+1)} \varepsilon E$ <u>and</u> $p = 2m + 1$, <u>then</u>

$$T_n(\phi_n, f) - f + \frac{(-1)^{m+1}}{n^{2m+1}} a_n \tilde{f}^{(2m+1)} = o\left(\frac{1}{n^{2m+1}}\right) .$$

We can give more precise equalities when we have better information about the processes or when we use the ω_m. Then we find in general cases the best asymptotic constants.

Marc Zamansky
Institut Henri Poincaré
11, rue Pierre et Marie Curie

75005--PARIS